

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

High stakes

The effects of cannabis
use on the brain's reward
pathways **PAGE 369**



GLOBAL WARMING

SOUTHERN COMFORT

For how long can the
Southern Ocean bale us out?

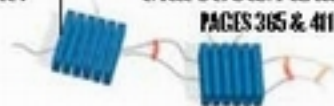
PAGE 346

FLEXIBLE ELECTRONICS

TRANSISTORS YOU CAN WEAR

Organic semiconductors
that stretch and heal

PAGES 365 & 411



CLIMATE SCIENCE

NEW LANGUAGE OF CARBON

In the right place, the
element's not the problem

PAGE 349

NATURE.COM/NATURE

17 November 2016 £10

Vol. 539, No. 7629



46 b

THIS WEEK



EDITORIALS

MARS Life in colony is not all work, rest and play **p.330**

WORLD VIEW A science in-tray for President Trump **p.331**

NEIGHBOURS Very early dinosaurs shared Earth **p.333**

Reality trumps rhetoric

As the shock of Donald Trump's election subsides, it is time for scientists and politicians alike to engage with core issues — from climate change and energy independence to social inequality.

There is a huge difference between campaigning and governing, as president-elect Donald Trump is surely realizing by now. It is impossible to know what direction the United States will take under Trump's stewardship, not least because his campaign was inconsistent, contradictory and so full of falsehood and evasion. People have not voted for his policies, because what are they? Americans have instead put their faith in an ideology, and rejected a political class and system that they feel, with some justification, has moved on without them.

In an Editorial last month, this journal argued that Trump was unsuitable for office (see *Nature* **538**, 289; 2016). His contrary approach to evidence, disrespect for those he disagrees with, and toxic attitudes to women and other groups have no place in a modern democracy. His election gives Trump the chance to prove the many people who shared that view mistaken. And that, we know for sure, is one thing Trump relishes. For those who opposed him, now is not the time to turn away from politics. There can be no normalizing or forgetting the malignant words and attitudes that Trump used on the campaign trail. But it is time to engage and to address the issues in a constructive manner.

A NATION DIVIDED

Many US citizens, including a fair number of scientists, might not like it, but Trump is a reflection of the United States today. He is a reminder of the deep schisms — economic as well as cultural — in American society and beyond. Academics, in particular, must break out of their cultural bubbles and work to understand the sentiments behind Trump's rise. There are elements of his agenda, including his attention to the plight of many working-class citizens who have missed out on the economic gains of the past 25 years, that truly merit attention. We need to better understand the causes and consequences of inequality, including how technology and globalization are reshaping the economic landscape.

In a World View on page 331, Daniel Sarewitz appeals to the pragmatist in Trump and suggests some science-based policies that he could support, if he wishes to make good on his promises to improve life in the American heartland.

Opponents should look for opportunities to compromise, too. Both Trump and Hillary Clinton called for increased spending on infrastructure, for instance, and this would be a good thing, if done wisely. As Barack Obama and Clinton said last week, the country — and indeed the world — needs Trump to succeed. But success in governance does not mean the implementation of a divisive agenda that caters to an angry minority, which is precisely what Trump's campaign was built on. The Democratic party still wields considerable power in the Senate, and it must use that power to oppose Trump's extreme agenda.

Much will depend on whom Trump appoints to key posts in the White House and the government. Trump's selection of Reince Priebus, chairman of the Republican National Committee, as his chief of staff may represent a pragmatic effort to mend ties with mainstream Republicans. But given the cheers coming from white nationalists, it

should come as no surprise that the ideologues will have their seat at the table as well. Trump has announced that his chief strategist will be none other than Stephen Bannon, who is executive chairman of the right-wing media site Breitbart News.

Trump has also selected a prominent climate sceptic, Myron Ebell, to head the transition team at the Environmental Protection Agency. Ebell, who is a policy director at the Competitive Enterprise Institute,

“Trump is unlikely to get everything he wants, even with a Republican majority.”

an advocacy group in Washington DC, has no scientific training, and has surely been appointed only to take a wrecking ball to the Obama administration's climate regulations. Here again, researchers must remain engaged and oppose all efforts to ignore the science on this or any other issue. *Nature* will play its part.

During his campaign, Trump advocated energy independence for the United States. But he has railed against subsidies for solar and wind power, and promised to tear up regulations that aim to trim the expansion and continuation of fossil-fuel use. The response to Trump will be shaped by how he adapts such firebrand messages to political and economic realities. Like President Obama, Trump is unlikely to get everything he wants, even with a Republican majority in both houses of Congress for at least the next two years. Indeed, less than a week after the election, he was already taking back some of his pledges. The “big, beautiful wall” he promised has been reduced to a fence in some areas. And after meeting Obama, Trump announced that he may seek to maintain parts of Obama's signature health-care reform.

Trump should drop his pantomime-villain act on climate change. If he does not, then, come January, he will be the only world leader who fails to acknowledge the threat for what it is: urgent, serious and demanding of mature and reasoned debate and action.

The world has made its decision on climate change. Action is too slow and too weak, but momentum is building. Opportunities and fortunes are being made. Trump the businessman must realize that the logical response is not to cry hoax and turn his back. The politician in Trump should do what he promised: reject political orthodoxy and listen to the US people. His Republican Party's position on climate change is rooted in the past. A March Gallup poll found that public concern about the problem had reached an eight-year high: some 64% of Americans, it suggested, worried a “great deal” or a “fair amount” about global warming. That is a lot more than the percentage who voted for President Trump.

There are nine more weeks until Trump swears to faithfully execute the office of president. As he does so, the world can only hope that Trump will respect evidence and expert advice. He has proved that he can stand up to withering criticism, from Democrats and Republicans alike. But can he change his mind in the face of hard facts? This is a true test for any leader, and Trump's legacy may well ride on the answer. ■

Maths for all

Explaining concepts to the ‘man in the street’ is a worthy goal, but is not always feasible.

David Hilbert was extremely absent-minded, extraordinarily brilliant and the most influential mathematician of the twentieth century. His reach continues today. Among other things, he popularized a common concept in the communication of science: the ‘man in the street’, whose understanding (or not) of a problem is commonly used as a benchmark for intelligibility.

At the 1900 International Congress of Mathematicians in Paris, Hilbert set out to list the most important open problems of the field for the new century. But he also emphasized communication. “A mathematical theory is not to be considered complete,” he said, “until you have made it so clear that you can explain it to the first man whom you meet on the street.” Hilbert attributed the saying to “an old French mathematician”.

That is unusually imprecise for a mathematician. So, to complete the theory, two intrepid maths historians set out to identify Hilbert’s elderly Gallic source. As they report in this month’s *Historia Mathematica*, they pursued a long paper trail, including a nineteenth-century letter published in *Nature*, and eventually succeeded (J. Barrow-Green and R. Siegmund-Schultze *Hist. Math.* **43**, 415–426; 2016).

It was indeed a Frenchman, one Joseph Diaz Gergonne (1771–1859), who first referred to the man in the street. In a letter dated 1825, Gergonne wrote that one has not said the last word on a theory until one has been able to explain it to a *passant dans la rue* — French for ‘passer-by in the street’. A year later, in a second letter, he went further. A formula or method that could not be explained to a passing stranger “does not deserve to see the light of day”.

That’s a big ask, because the man (or woman) in the street sets the bar high. As most readers of *Nature* will know well, explaining your research to a non-expert is often challenging. Certainly, Hilbert acknowledged that the goal was an exaggeration. A more realistic

one might be that attributed to Albert Einstein, who said (perhaps apocryphally) that things should be made as simple as possible, but no simpler.

That tension must have been on the mind of Nobel prize committee member Thors Hans Hansson in October. At a press conference, he was tasked with explaining the motivation for awarding this year’s physics prize to David Thouless, Duncan Haldane and Michael Kosterlitz, who brought topology into physics in unexpected ways. In an attempt to illustrate the concepts involved, Hansson produced a bagel, a pretzel and a cinnamon bun from a bag.

“The man (or woman) in the street sets the bar high.”

Topology is a relatively young branch of maths: when Hilbert made his list of open problems, he included few problems from topology. The field studies continuous deformations. The explanation goes that bagels and pretzels have different topologies because turning the former into the latter involves tearing the dough apart — not a continuous deformation.

There is no word yet from the street on whether Hansson was successful in his explanation. But Gergonne would surely have approved: he studied projective geometry, which involves shapes with exotic topologies, such as surfaces that have only one side. He would also, surely, be astonished that his 1825 passer-by continues to pass by the labs and lecture rooms of the twenty-first century, ever-ready to test his or her comprehension.

As for Einstein, Hilbert at one point raced him to the formulation of the general theory of relativity, only to later graciously concede. Hilbert is also credited with the dubious assertion that “Every boy in the streets of Göttingen understands more about four-dimensional geometry than Einstein”, which seems to be taking the principle of the passer-by as arbiter a little too far. (In fairness, he then went on to admit: “Yet, in spite of that, Einstein did the work, and not the mathematicians.”)

The new historical analysis solves one of the questions posed by Hilbert’s famous speech, but plenty remain. The celebrated 23 Hilbert problems inspired researchers for decades; some remain unsolved. Luckily, the passer-by can wait. When a solution appears, he or she will surely be ready. ■

Home from home

We cannot look to the stars to solve our planet’s problems.

World events this year have been thought-provoking. One thought present in many minds might be: how are the plans for a Moon base or a Mars colony coming along? So, some much needed good news! US engineers have published their design for a nuclear reactor that could power a permanent off-planet settlement for 15 years (K. J. Schillo *et al. Ann. Nucl. Energy* **96**, 307–312; 2016).

In 2009 (and how long ago that seems now), NASA identified such a reactor design as essential if astronauts were to visit and spend time on Mars. It runs (but these are mere details) on low-enriched uranium ceramic-metal fuel and uses supercritical carbon dioxide — cleverly extracted from the Martian atmosphere — as a coolant. With a reliable source of electricity, a Mars colony could then apply itself to what NASA calls in-situ resource utilization, and what everybody who saw the 2015 film *The Martian* will think of as growing potatoes in poo.

Don’t pack your bags for this off-world utopia just yet. For what kind of society will such an isolated outpost create? A series of papers in the journal *Space Policy* speculate on human affairs on the red planet, and reach some depressingly familiar conclusions. The costs of getting to

and living on the planet suggest the need for heavy corporate involvement, which could establish a conflict between those who want Mars to remain, well, Martian, and those who would develop and exploit its resources. Explorers must be able to plant a flag and claim territory (contrary to existing space law) to justify the trip — but, without restrictions, a free-for-all on the Mars landscape could ruin it for everyone.

The offered solution, naturally, is to divide Mars into exclusive economic zones, separated by a string of protected planetary parks. These would be safeguarded, naturally, by rules: no littering with spacecraft parts, and no walking or driving except on designated tracks.

Rules create a problem for colonists: whose interests do they serve, and who gets to decide them? A central Martian authority is a non-starter — attempts to impose a lunar government have stymied progress on a parallel Moon treaty — so some kind of tribunal system will be necessary, with appointments no doubt squabbled over by existing and new powers. Finally, there is the question of what happens should the colonists decide that they are sick of being told what to do by their parent planet. How should Earth respond to a Martian rebellion or conflict? Would we intervene? Would we take refugees?

One ‘pragmatic’ answer to all this likely division, researchers suggest, is the deliberate development of a new Mars religion, especially for those born on Mars. This could emphasize the sense and purpose of the mission, and help to justify the difficult living conditions.

The picture of Mars painted by these discussions, in other words, is a planet divided by politics, culture, religion, economics and inequality. Sound like anywhere you know? ■

MICHAEL TEMCHINE



Science and innovation policies for Donald Trump

Daniel Sarewitz says the new US president will need to promote creative science to benefit the disaffected millions who voted for him.

America's 'rust belt' states, which gave Barack Obama their votes in 2008 and 2012, have now sent Donald Trump to the White House. It's a shocking reversal that testifies to disaffection on many fronts. But it's also a long-term political consequence of bipartisan trade, economic and innovation policies that have focused on growth and competitiveness while largely neglecting the negative impacts of technological change and globalization on quality of life in the American heartland.

Trump built his campaign on the promise to help a middle class struggling from such dislocations to regain its footing. If he is serious, he will need to mobilize smart and creative science and innovation policies. Here are some.

One rare bipartisan initiative of the past few years has been Manufacturing USA, a network of nine regional centres focused on specific technological challenges such as 3D printing, photonics and functional fabrics. Each centre brings together private, government and university partners to help small firms to scale up and commercialize new technologies.

The Trump administration should significantly build on this programme and integrate it with the Manufacturing Extension Partnership, which provides small manufacturers with technical assistance. The overall goal should be regional innovation networks that draw from local universities, talent and resources to build new industries, expand manufacturing and create jobs.

Initially, these and other programmes could be funded by Trump's plan to repatriate profits, amounting to up to US\$2 trillion, that US corporations have moved offshore. Trump wants to bring the money back to the United States by offering a one-time repatriation tax rate of 10% as an incentive. The resulting \$200 billion or so of revenue is the equivalent of nearly three years of total non-defence government research and development spending.

Repatriation of offshore pharmaceutical-company profits alone could generate \$40 billion of revenue. This money could be spent on high-priority health-care problems relevant to Trump's agenda. First, an organization similar to the Defense Advanced Research Projects Agency is needed to bring academic, private-sector and government research and development capacities together to pursue high-risk, high-gain solutions to health problems such as obesity, diabetes and substance abuse, which are especially acute in economically depressed areas of the nation. Second, a national database should be created to integrate biomedical research data with electronic health records. Such a system is a prerequisite for a 'learning' health-care system that can use the huge amount of information generated both at the bench and in the clinic to deliver better health outcomes. Only government can do this.

Trump is publicly sceptical about the seriousness of climate change,

and is committed to the coal industry. Yet climate change offers him the opportunity to steal one of the Democrats' key issues while advancing the nation's economic interests and energy security. As a supporter of nuclear energy, he could increase funding for research, development and demonstration of next-generation reactors — on which the United States lags far behind China and Russia — and reopen the licensing process for the Yucca Mountain waste-storage facility.

Meanwhile, Trump could reignite efforts to develop coal plants that co-fire biomass and capture carbon dioxide — thus protecting jobs, creating economic opportunities for rural communities that provide fuels such as timber residues, and producing meaningful emissions reductions. Power companies could in turn sell desperately needed CO₂ to the oil industry for advanced oil recovery — stimulating private-sector

investment in infrastructure for CO₂ transport, while returning more CO₂ to the ground. Trump would barely have to soften his stance — just enough to acknowledge the pragmatism in measures that both benefit the economy and reduce the threat posed by greenhouse gases.

Once committed to an increasingly clean electricity grid, a Trump government could further invest in a national network of electric-vehicle recharging stations — thus solidifying the nation's competitive advantage in this key technology, and taking a giant step towards the elusive goal of a low-carbon transport infrastructure.

There are other ways for the anti-regulatory president-elect to create growth while protecting the environment. The US chemicals industry accounts for more than 10% of the nation's manufacturing revenue and nearly 800,000 jobs

in a highly competitive global market. A Trump regime could adopt the proven model of the Massachusetts Toxics Use Reduction Act, in which toxic chemicals are not directly regulated and banned, but firms work with government and academia towards assessing and adopting safer alternative chemicals and processes. Such an approach would help to protect the health and jobs of the manufacturing workers and communities Trump is concerned about, while stimulating innovation.

There is much — much — to worry about with Trump's election, starting with his apparent intolerance of fundamental tenets of civil society, such as rule of law and freedom of speech. But it's also true that, as a political outsider, he is less beholden to constituencies that would limit the political options of more-mainstream candidates on both sides. As such, his presidency at least has the potential to open up productive new policy pathways. We will soon find out if he is serious about doing so. ■

Daniel Sarewitz is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC. e-mail: daniel.sarewitz@asu.edu

**TRUMP COULD
REIGNITE
EFFORTS TO DEVELOP
COAL PLANTS
THAT CO-FIRE
BIOMASS
AND CAPTURE
CARBON DIOXIDE.**

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PHOTONICS

Nanolasers for precision imaging

Researchers have turned nanowires into tiny lasers that could one day be used in high-resolution microscopy.

The resolution of optical microscopes is normally constrained by the wavelength of light. To overcome this 'diffraction' limit, a team led by Seok Hyun Yun at Massachusetts General Hospital in Cambridge made nanowires of lead iodide perovskite that were 3–7 micrometres long and just 300–500 nanometres wide. When the team scanned the nanoparticles with a laser, they emitted light. The narrow width of the nanowires allowed their positions to be tracked with resolutions roughly five times better than the diffraction-limited resolution.

Such nanolasers could one day be placed inside cells and yield super-resolution images from deep within tissue samples, the authors say.

Phys. Rev. Lett. 117, 193902 (2016)

GEOCHEMISTRY

Plants take up more carbon

Enhanced plant growth over the past decade seems to have slowed the build-up of carbon dioxide in the atmosphere.

Despite the rise in CO₂ emissions resulting from human activity, atmospheric CO₂ levels have grown relatively little since 2002. To find out why, Trevor Keenan at Lawrence Berkeley National Laboratory in California and his colleagues used ground observations, satellite data and vegetation models to quantify changes

in CO₂ uptake and release by terrestrial plants worldwide. They found that increased photosynthesis and plant 'greening' have boosted the amount of carbon stored on land. Reduced plant respiration due to the recent slowdown in the rate of global warming also seems to have increased this carbon sink.

However, terrestrial carbon stocks will not offset the accumulation of heat-trapping gases in the atmosphere if emissions continue to grow, the scientists caution.

Nature Commun. 7, 13428 (2016)

ANIMAL BEHAVIOUR

Plastic smells good to marine birds

Marine animals often mistake plastic pollution for food, possibly because of its smell.

Marine grazers such as krill consume microscopic organisms called phytoplankton, which release dimethyl sulfide (DMS). Some of krill's seabird predators, such as petrels and shearwaters, sniff out this chemical to find the grazers. Matthew Savoca and Gabrielle Nevitt at the University of California, Davis, and their colleagues attached beads of the three most

common types of ocean plastic to buoys off the coast of California for three weeks. They detected DMS emitted from every seawater-exposed sample, but not from unexposed plastic.

Data from other studies showed that predators that were most responsive to DMS also ingested the most plastic, which can poison animals and block their digestive systems (pictured is an albatross with plastic in its stomach).

Sci. Adv. 2, e1600395 (2016)



PAUL & PAVEENA MCKENZIE/GETTY

CANCER IMMUNOTHERAPY

Gene-edited cells fight cancer

Immune cells engineered to target cancer could provide an off-the-shelf therapy if results in mice can be replicated in people.

Previous work has shown that engineered T cells called CAR-T cells can be taken from a patient and modified to kill some types of tumour, but the supply can be limited. To make a more reliable, universal source of donor cells, Carl

June, Yangbing Zhao and their colleagues at the University of Pennsylvania in Philadelphia used the genome-editing tool CRISPR-Cas9 to tweak human CAR-T cells. They disabled two proteins that can trigger immune rejection of donor T cells, resulting in a reduced reaction from the immune system when the cells were used in mice.

The team also disrupted a protein called PD-1, which normally holds immune responses in check. The cells slowed the growth of human tumours implanted into mice.

Further testing of immune responses should be carried out in non-human primates, the authors say.

Clin. Cancer Res. <http://doi.org/bsvc> (2016)

NEUROSCIENCE

Brain circuit for tickling found

Researchers have pinpointed a brain circuit in rats that fires when the animals are tickled.

Rats are known to chirp or 'laugh' when tickled. To find the neural mechanism, Michael Brecht and Shimpei Ishiyama at the Humboldt University of Berlin inserted electrodes into the somatosensory area of the rat brain, which responds to touch on the skin. Certain neurons in this region fired intensely when the rats were tickled on their bellies, but responded less to back-tickling and hardly at all to tail-tickling.

By stimulating these neurons with the electrodes, the researchers caused the rats to chirp in the same way they do when tickled.

Science 354, 756–759 (2016)

CARDIOVASCULAR BIOLOGY

Food chemical protects the heart

A compound found naturally in legumes, grains and mature cheese helps rodents to maintain a healthy heart.

The chemical, spermidine, extends the lifespans of yeast, flies and worms by promoting autophagy, a cellular process that degrades and recycles old components of the cell and toxic products. Frank Madeo at the University of Graz in Austria and his collaborators found that aged mice that regularly drank water containing spermidine had improved heart function compared to untreated mice of the same age, and that the compound lowered blood pressure in rats prone to developing heart failure. Spermidine reversed various age-related mechanical and metabolic changes in the heart

cells of treated animals.

The chemical seemed to exert its effects through autophagy, because it did not benefit mice with a genetic defect that impairs this process in heart-muscle cells.

Nature Med. <http://dx.doi.org/10.1038/nm.4222> (2016)

MATERIALS

Elastic polymer mimics muscles

A lightweight elastomer can stretch and contract when stimulated by low voltages, making it a promising material for artificial muscles.

Dielectric elastomers exhibit reversible physical deformations when stimulated by electricity, but previous attempts to make artificial muscles from such materials required cumbersome braces to prevent rupture during long stretches. To avoid using braces, Andrey Dobrynin at the University of Akron in Ohio, Sergei Sheiko at the University of North Carolina at Chapel Hill and their team developed a dielectric elastomer by linking together polymer strands in a bottlebrush-like structure. This resulted in a stronger and more stretchable material with a lower risk of rupture than current materials.

The synthetic elastomers could eventually have applications in fields including soft robotics, the authors say.

Adv. Mater. <http://doi.org/f3s534> (2016)

INFECTION

Red squirrels get leprosy

Two species of bacteria that cause leprosy have been found in red squirrels in Britain and Ireland — a surprising discovery because only

primates and armadillos were thought to get the disease.

Stewart Cole at the Swiss Federal Institute of Technology in Lausanne, Anna Meredith at the University of Edinburgh, UK, and their team analysed tissue samples from 110 dead red squirrels (*Sciurus vulgaris*). Thirteen showed characteristic signs of infection such as skin lesions and swelling of the snout. The authors found leprosy-causing bacteria — *Mycobacterium leprae* or *Mycobacterium lepromatosis* — in all visibly ill animals (squirrel with leprosy on ear and snout pictured) and in about one-fifth of symptom-free ones.

This could represent another threat to red squirrels in the United Kingdom and Ireland, which are already affected by other infections and by competition from grey squirrels introduced from North America.

Science 354, 743–746 (2016)

PALAEOONTOLOGY

Early dinosaurs emerged gradually

Dinosaurs and their precursors lived side by side in the earliest stages of dinosaur evolution.

Max Langer at the University of São Paulo in Brazil and his colleagues explored some of the oldest-known rocks that contain dinosaur fossils, aged between 237 million and 228 million years old. The team unearthed a new species of an early 'lizard-hipped' dinosaur (*Buriolestes schultzi*; pictured right), which co-existed with a non-dinosaur relative, *Ixalerpeton polesinensis* (left). The non-dinosaur had skull and bone characteristics that resemble those of later dinosaurs, showing where key parts of dinosaur anatomy came from. The dinosaur's



teeth suggested that it ate small animals, unlike later species in its class that fed only on plants.

The findings suggest that dinosaurs did not rapidly replace their predecessors during their early evolution.

Curr. Biol. <http://doi.org/bsww> (2016)

ARCHAEOLOGY

Old DNA reveals fishing habits

Analysing ancient DNA from the tropics is difficult because DNA breaks down fairly rapidly in heat, but a team has managed to tease out some of the species represented by hundreds of 100–300-year-old fish bones from Madagascar.

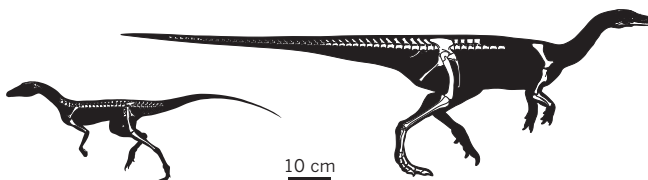
Short 'barcode' DNA sequences from a region of the genome that tends to vary between taxonomic groups can be used to identify fragmented animal bones discovered at archaeological sites. A team led by Alicia Grealy at Curtin University in Perth, Australia, applied this method to 887 fish bones from a site on the west coast of Madagascar that was occupied by humans in the eighteenth and nineteenth centuries. The authors were able to identify 23 fish families — 10 of which could be narrowed down, with some confidence, to a specific species — including two kinds of shark and various coral-reef-dwelling fish.

J. Archaeol. Sci. 75, 82–88 (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



EVENTS

Bird flu in Europe

A highly pathogenic strain of avian influenza virus called H5N8 has been identified in several European countries, threatening the region's poultry industry, worth €30 billion (US\$32 billion) a year. On 12 November, authorities ordered the slaughter of 30,000 chickens on a farm in northern Germany as a containment measure. Scientists believe that the strain was imported from Asia by migrating birds. The last epidemic of H5N8 in Europe was in November 2014; more than 336,000 birds were culled. H5N8 is not known to infect humans.

Computing cancer

A genomics powerhouse will team up with a supercomputer to unpick the tangled cellular networks that can cause resistance to cancer drugs. On 10 November, the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, and IBM Watson Health announced a five-year, US\$50-million study that uses IBM's Watson artificial-intelligence system to analyse the genomes of thousands of drug-resistant tumours. The goal is to find patterns that could allow clinicians to predict drug sensitivity and resistance. The data are to be shared with the scientific community.

Stem cells on offer

Public prosecutors in Italy are investigating whether disgraced entrepreneur Davide Vannoni — who was convicted on criminal charges last year for administering unproven stem-cell therapies — is offering his procedures again. The terms of his conviction required him to refrain from performing procedures in Italy, and possibly from doing

so abroad. However, patient groups on social media posted comments suggesting that the procedure is available in Georgia. The technique involves modifying stem cells from the bone marrow of a patient or relative and injecting them back into the patient. In October, a patient came forward with a detailed account of his treatment, prompting prosecutors to take action. See page 340 for more.

PUBLISHING

Open peer review

A trial by *Nature Communications* shows that many scientists are willing to try open peer review. Around 60% of the journal's authors

so far this year have agreed to have their reviews published, so the journal will continue to offer the option — but will not make it mandatory, it said on 10 November. See page 343 for more.

Wellcome boost

The Wellcome Trust in London, one of the world's largest medical charities, launched its open-access publishing platform on 15 November. *Wellcome Open Research* allows Wellcome grant recipients to publish a variety of findings — including conventional articles, software, study protocols and living data sets — within days of submission. Publication is followed by transparent peer

on 14 November. The flattening is largely the result of a drop in coal consumption in the United States and China, the world's largest greenhouse-gas emitters. If warming is to be kept below 2 °C, global CO₂ emissions must soon start to decline by around 3% per year, scientists reminded the United Nations climate meeting now under way in Marrakesh, Morocco.



PATRICK PLEUL/AFP/GETTY

Carbon dioxide emissions stay stable

review in which authors choose their referees. The funder hopes that its initiative will change the way research is assessed, and that other grant-givers will follow its model.

RESEARCH

LIGO restart

After a ten-month shutdown for upgrades, the Laser Interferometer Gravitational-wave Observatory (LIGO) is ready to make discoveries again. In February, researchers stunned the world when they announced the detection of gravitational waves from a collision between distant black holes, and a second detection was confirmed in June. LIGO's two 8-kilometre-long

ESA/DENMAN

interferometers, one in Louisiana and the other in Washington state, began a new 'engineering run' on 14 November and expect to start science runs in December. Upgrades to the detectors should enable the observatory to scan a significantly larger volume of the Universe than before.

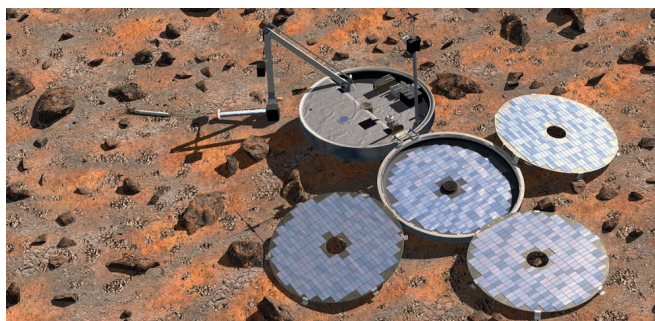
SPACE

Space debris

A rocket stage from a Chinese launch apparently crashed near a jade mine in northern Myanmar on 10 November. That morning, a Long March 11 rocket lifted off from the Jiuquan launch centre in Inner Mongolia; minutes later, it would have passed over northern Myanmar. Local media reported the sound of an explosion and a stench in the air when the object fell. The rocket successfully put five satellites into orbit, including one meant to test celestial navigation using X-rays emitted by pulsars.

Beagle 2 worked

Beagle 2 (pictured, artist's impression), which lost touch with Earth as it landed on Mars on Christmas Day 2003, came extremely close to success, mission scientists announced on 11 November. By comparing images of the lander taken by NASA's Mars



Reconnaissance Orbiter with simulations of how sunlight would reflect off the probe in various configurations, they found that Beagle 2 probably opened at least three of its four solar panels. The British-built lander, operated by the European Space Agency, may even have collected data, but the last panel's failure to fully open would have blocked transmission from the craft's antenna, says the UK team behind the analysis, which is not yet published.

POLITICS

President Trump

The election of Republican businessman and reality-television star Donald Trump as the next president of the United States on 8 November sent shock waves around the world. Trump's views on women and minority groups, his stance on climate change and nuclear weapons, and his contradictory approach to evidence have worried many

— including scientists. During his successful campaign to defeat the Democratic candidate Hillary Clinton, Trump pledged to remove the United States from the 2015 Paris climate-change agreement and to dismantle the deal between six world powers and Iran to limit that country's nuclear capabilities. See pages 329, 331 and 337 for more.

State initiatives

Americans voted on a host of state-level initiatives on 8 November, some science-related. Voters in Washington state rejected a carbon tax, with its backers saying they will bring a reworked proposal to the Washington legislature next year. The people of Montana rejected a plan that would have awarded up to US\$200 million in grants to biomedical scientists over 10 years. And voters in southern Florida were split on whether to back a plan to release genetically modified mosquitoes to fight diseases such as the Zika virus. It was

COMING UP

17 NOVEMBER

Europe's next four Galileo navigation satellites are scheduled to launch on an Ariane 5 rocket from French Guiana, bringing the number of satellites in the constellation to 18. go.nature.com/2fmfxfs

19 NOVEMBER

The United States launches GOES-R, its first next-generation geostationary weather satellite, from Cape Canaveral, Florida. www.goes-r.gov

23 NOVEMBER

The UK government makes its first major financial statement since the vote to leave the European Union in June.

rejected by the town of Key Haven, where the field trial would take place, but voters in surrounding Monroe County approved it. Ultimate approval rests with the Florida Keys Mosquito Control District, which has yet to decide on the matter.

POLITICS

Science adviser

Canada will issue an open call for nominations for the position of national science adviser, its science minister announced on 10 November. The Canadian public, researchers and institutions will be asked to suggest candidates for the post, which Liberal Prime Minister Justin Trudeau promised to bring back after being elected in 2015. The post had been scrapped in 2008 by his Conservative predecessor, Stephen Harper. Trudeau has also vowed to usher in an era of evidence-based policymaking.

NATURE.COM

For daily news updates see:

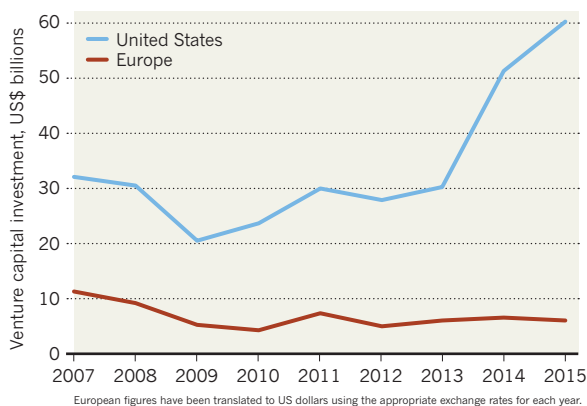
www.nature.com/news

TREND WATCH

Europe's research commissioner, Carlos Moedas, has launched an effort to boost venture capital (VC). Last year, investors put €5.3 billion (then US\$5.9 billion) into firms in Europe; US investment was \$60.1 billion. To help close the gap, Moedas said on 8 November that the European Commission would offer €400 million for new independently managed VC funds, provided they raised triple that amount from private investors. The funds would go to other VC funds, rather than company stocks.

EUROPE'S VENTURE-FUNDING GAP

Europe's venture-capital investment severely lags behind that in the United States.



SOURCE: EUROPE INVEST AND MONEYTREE

NEWS IN FOCUS

HEALTH Italian stem-cell entrepreneur under fresh investigation **p.340**



EMOJICON The science-themed icons coming to a phone near you **p.341**

RWANDA Declines in foreign aid threaten science success story **p.342**

CLIMATE How much longer will the Southern Ocean delay global warming? **p.346**

ZACH GIBSON/GETTY



US president-elect Donald Trump (right) surveys the National Mall in Washington DC.

POLITICS

The Trump experiment

Researchers struggle to predict US president-elect's impact on science.

BY SARA REARDON, JEFF TOLLEFSON, ALEXANDRA WITZE & LAUREN MORELLO

The long campaign for the White House is over — but incoming US president Donald Trump's work is just starting. With about two months to go before his inauguration on 20 January, he and his staff are busy vetting candidates for top government jobs and clarifying his agenda for governing.

Some scientists have expressed fear about how Trump's presidency will affect research in the United States. The president-elect has questioned the science underlying climate change and has linked autism to childhood

vaccinations; the vice-president-elect, Indiana governor Mike Pence, does not believe in evolution or that human activities have caused climate change. Still, some science advocates caution against a rush to judgement about how the Trump administration will approach science and research issues.

"The verdict remains out," says Tobin Smith, vice-president for policy at the Association of American Universities in Washington DC. "There are many people who have been strong supporters of science you might not have been expecting," Smith says that a prime example is Newt Gingrich, the former Republican congressman for Georgia, who is rumoured to be

up for a top job in Trump's administration.

As speaker of the US House of Representatives in the 1990s, Gingrich supported a plan to double the National Institutes of Health (NIH) budget over ten years. Since leaving Congress, he has advocated significant spending hikes for the National Science Foundation and other science-funding agencies.

But it's hard to draw any conclusions about Trump's views on science, given his limited comments on such issues, other policy specialists caution. "He speaks positively of innovation, but 'innovation' is a big word," says Kevin Wilson, director of public policy and media relations at the American Society for Cell



► Biology in Bethesda, Maryland. “You can drive a truck through innovation. We don’t know what that means.”

Here, *Nature* looks at key science issues that Trump will confront during his first few months in office.

BIOMEDICAL SCIENCE

Trump has said that one of his priorities after taking office will be to reverse several executive orders that President Barack Obama, a Democrat, issued on topics ranging from climate change to immigration. Some biomedical researchers are worried that Trump will cancel an Obama order that authorized experiments with human embryonic stem cells.

“It’s something very tangible,” says Wilson. “He could do away with it on day one.”

There’s precedent for such actions: Obama’s March 2009 executive order on stem cells reversed limits put in place by his predecessor, Republican George W. Bush.

And Pence opposed Obama’s decision to authorize research with human embryonic stem cells. “It is morally wrong to create human life to destroy it for research,” he wrote in a March 2009 newspaper commentary.

Overall, however, Trump has said little about biomedical science — aside from an oft-quoted 2015 radio interview in which he called the NIH “terrible”.

Mary Woolley, president of the advocacy group Research!America in Arlington, Virginia, worries that biomedical science will not be a priority for Trump. “A lot of it is not really controversial,” she says. “We tend in this country to take progress for granted.”

CLIMATE CHANGE

If Trump keeps his promises, the United States will reverse course on global warming. The president-elect has blasted the US Environmental Protection Agency (EPA) and said that he will repeal Obama’s climate regulations. And Myron Ebell, a prominent climate sceptic who directs energy and global-warming policy at the Competitive Enterprise Institute in Washington DC, is leading Trump’s transition team for the EPA.

“I take Trump at his word,” says Jeffrey Holmstead, an attorney at the firm Bracewell in Washington DC who worked at the EPA under George W. Bush. “And I think they won’t



SCIENTISTS SPEAK

How researchers reacted to the election of Donald Trump.

“I don’t actually know a Trump supporter who I could talk to about the election. How can I reach the public if I’m only speaking to my own circle?”

Peter Peregrine, anthropologist at Lawrence University in Wisconsin and the Santa Fe Institute in New Mexico.

“I feel sad for the United States and its scientists, and would like to welcome good scientists to work and live in China.”

Yi Rao, neuroscientist at Peking University.

“Political events do not and cannot change the reality of climate change.”

Philip Duffy, president of the Woods Hole Research Center in Falmouth, Massachusetts.

“As a Canadian working at a US university, a move back to Canada will be something I’ll be looking into.”

Murray Rudd, studies environmental economics and policy at Emory University in Atlanta, Georgia.

For more reactions, see go.nature.com/2fu4crp



have any difficulties.”

Trump’s first target will probably be the Clean Power Plan, Obama’s regulations to reduce greenhouse-gas emissions from power plants, which roughly two dozen states are challenging in court. The case is expected to reach the Supreme Court as early as next year. By then, Trump may have filled the court’s current vacancy with a conservative justice. That would put the climate regulations in jeopardy.

But it would be easy enough for the Trump administration to just revoke the Clean Power Plan on its own, Holmstead says.

Trump’s vow to pull the United States out of the Paris climate accord — a process that could take four years to complete — is still sinking in at the United Nations climate talks in Marrakesh, Morocco. Delegates there are busy hashing out a plan to implement the Paris agreement. “We’re sort of in the denial stage,” says Jake Schmidt, international programme director at the Natural Resources Defense Council in New York City. “I suspect there will be some disappointment and anger starting to bubble up in the next couple of days.”

Already, many nations are looking to China for leadership on climate. The country leads the world in renewable-energy investment

because it views clean energy as a necessity and an opportunity, says Andrew Steer, president of the World Resources Institute, an environmental think tank in Washington DC. He hopes that Trump will come to see tough climate-change policies as a tool to ensure that the United States remains competitive in the development of energy technology.

SPACE

Trump himself has said little about space policy, but astronaut Eileen Collins — the first woman to command a space shuttle — spoke at the Republican national convention in July. Collins called for the United States to reassert its leadership in space exploration.

In October, two of Trump’s campaign advisers wrote a pair of commentaries in SpaceNews laying out possible directions for space policy under a new president. The articles argued that NASA should focus more on deep-space exploration and less on what they called “politically correct environmental monitoring”.

NASA’s Earth-observing missions account for more than one-third of the agency’s science budget, an expense that has come under fire from congressional Republicans. “It’s conceivable that the Trump White House could go after NASA Earth science,” says John Logsdon, former director of the Space Policy Institute at the George Washington University in Washington DC.

The Trump advisers also argued for more public-private partnerships in civilian space. Such efforts are already under way with private companies now ferrying US cargo, and soon US astronauts, to the International Space Station.

Casey Dreier, director of space policy for the Planetary Society in Pasadena, California, says that space is likely to be a low priority for Trump during his first 100 days as president.



TOP NEWS

- What scientists should focus on — and fear — under Trump go.nature.com/2evrfcd
- AI science search engines reach out go.nature.com/2ew0f15
- Playful rats reveal brain region that drives ticklishness go.nature.com/2gcwpty

NATURE PODCAST

Pluto’s icy ‘heart’; your brain on cannabis; and all about living on Mars nature.com/nature/podcast



POLITICS

Why the polls missed Trump

Pollsters lament failure to foresee outcome of US election.

BY RAMIN SKIBBA

What went wrong? That's the question many political pollsters in the United States are asking themselves in the aftermath of the 8 November presidential election. Republican candidate Donald Trump won in an electoral landslide, but for months most polls forecast a victory for his Democratic opponent, Hillary Clinton.

Many types of poll, including randomized telephone polls and online polls that people opt into, indicated a tightening of the gap between the two candidates in the weeks leading up to the election — but still pointed to a Clinton win. “The industry is definitely going to be spending a lot of time doing some soul-searching about what happened,” says Chris Jackson, head of US public polling at Ipsos, a global market-research and polling firm based in Paris.

The most recent national polls — including those conducted by ABC News/*Washington Post*, Ipsos, YouGov and Fox News — all estimated a Clinton lead of 3–4% over Trump. Yet as the last votes are being counted, Clinton leads the popular vote by a razor-thin margin: just 0.2%. The majority of states have tipped for Trump, giving him their valuable electoral-college votes and ensuring his victory. These include several Midwestern states that Clinton was expected to win.

Poll aggregators such as FiveThirtyEight nonetheless forecast Clinton's chances of victory at 71% or higher. This dramatic polling failure could have been due to poorly assessed voters, people misreporting their voting intentions, or pollsters inadequately surveying some segments of the population.

“It's a big surprise that such a wide variety of polls using such a wide variety of methodologies have all the errors fall in the same direction,” says Claudia Deane, vice-president of research at the Pew Research Center in Washington DC.

The University of Southern California Dornsife/*Los Angeles Times* presidential

election poll, which included an online panel of nearly 3,000 people, was the only major national poll to forecast a Trump lead days before the election. “But we're not sure we were right either,” says Jill Darling, survey director at the university's Center for Economic and Social Research in Los Angeles. She notes that Trump did not defeat Clinton by 3%, as her group's most recent poll predicted.

With each election, pollsters have a harder time reaching people. Now that Americans have fewer landlines and more mobile phones with caller ID, they don't respond to calls from unfamiliar numbers. Online surveys also struggle to recruit participants. A poll generally needs at least 1,000 participants who are representative of the general population with respect to gender, race, education, income level and geographic distribution to produce statistically significant results.

Pollsters strive to assess not just who supports whom, but also who will be likely to vote. This year, 119 million people cast ballots, accounting for 55.6% of registered voters, according to Michael McDonald, a political scientist at the University of Florida in Gainesville. That is the lowest percentage since 2000.

There were also more undecided voters this year than in previous presidential elections. Such voters may be under-represented in polls, yet tilt towards one candidate, Darling says. Only 53% of poll respondents disclosed who they would vote for, lower than the 70% in earlier elections, she adds. And people overestimate their own likelihood of voting.

“It seems like Trump voters were more enthusiastic about turnout and less enthusiastic about responding to polls. That's a deadly combination,” says Andrew Gelman, a statistician and political scientist at Columbia University in New York City.

Polling experts in Britain conducted a formal inquiry following polling failures in last year's general election, when polls underestimated the turnout of older, Conservative voters. Now, in the United States, the American Association for Public Opinion Research has already named an ad-hoc committee to dig into the data and conduct a post-mortem on the election polls. They aim to produce findings by next May, Deane says. ■

See go.nature.com/2f9hpeo for a longer version of this story.



Falcon9 rockets built and launched by SpaceX carry cargo to the International Space Station.

SPACEX

Dreier will be watching whether the new Congress cuts government spending. “If that's the case, NASA will be impacted by that along with every federal agency,” he says.

IMMIGRATION

Trump reinvigorated the national debate on immigration with his campaign pledges to build a wall along the US border with Mexico and to temporarily ban Muslims from entering the United States.

“Our hope is that the rhetoric of the election was only a façade for something hopeful that's going to be more pragmatic and engaging communities,” says Carl Saab, a neuroscientist at Brown University in Providence, Rhode Island, and the former president of the Society for Arab Neuroscientists.

Trump has variously said that the ban would apply to all Muslims and to anyone from “nations tied to Islamic terror”, drawing vigorous criticism from civil-liberties groups that say such a policy would violate the US Constitution. He has also proposed deporting more people who are in the United States illegally, which could include those who came to the country as children.

Some researchers worry that such policies would threaten US research dominance. About 5% of US university students come from other countries, including more than 380,000 people studying science, engineering, technology or mathematics.

“The rhetoric that Mr Trump ran under has frightened lots of immigrants,” says Benjamin Corb, director of public affairs for the American Society for Biochemistry and Molecular Biology in Rockville, Maryland. “I certainly hope that we don't end up losing some brilliant minds as a result of some near-sighted policies.” ■



Davide Vannoni during court proceedings against his Stamina Foundation in 2014.

REGENERATIVE MEDICINE

Stem-cell scandal gets fresh scrutiny

A treatment banned in Italy may have resurfaced abroad.

BY ALISON ABBOTT

Public prosecutors in Turin, Italy, are investigating whether disgraced stem-cell entrepreneur Davide Vannoni — convicted on criminal charges last year for administering unproven stem-cell therapies in Italy — is offering his treatments again, this time in eastern Europe.

In March 2015, Vannoni was convicted on charges of conspiracy and fraud related to his treatments, which had been declared dangerous by the Italian Health Authority (AIFA). His case was a cause célèbre among Italian scientists, who fought for many years to stop him administering stem cells to patients through his Stamina Foundation (see *Nature* **518**, 455; 2015).

Vannoni was sentenced to 22 months in prison, but the sentence was suspended in a plea bargain drawn up by prosecutor Raffaele Guariniello, who said that the terms required Vannoni to refrain from organizing further therapies — either in Italy or abroad. Soon after the plea bargain, patient groups on social media posted comments that the stem-cell treatment was available once more, in Georgia. And in late October, a patient came forward with a detailed account of his treatment. Prosecutors told *Nature* that this has prompted them to take action.

Vannoni did not respond to requests for comment, forwarded through his lawyer, Liborio Cataliotta. But Cataliotta told *Nature*

that, although he did not know whether Vannoni was continuing therapies in Georgia, he thought there were “technical defects” in what Guariniello had said about the terms of the plea bargain. The prosecutor had no authority to rule on what might be carried out in countries where such stem-cell therapies are allowed, he said.

Guariniello has since retired, but another prosecutor in Turin, Vincenzo Pacileo, told *Nature* that the deferred sentence hinges on the premise that Vannoni “gives up” his stem-cell activities, a phrase that Pacileo says implies both in Italy and abroad. However, a senior Italian legal expert who has read the 46-page sentence, but does not wish to be named, says that the premise is legally ambiguous. Nowhere does the document specify that this applies to activities abroad, he says.

“It is really disappointing that science has not been able to put an end to this,” says Luca Pani, a former director-general of AIFA. “A patient is a patient wherever they are treated, and I worry for them and their families.” Pani tried to close down Stamina’s operations in Brescia in 2012, but desperate patients got court orders allowing them to be treated. It took two years before magistrates in Turin ordered the confiscation of equipment and materials from the Stamina lab.

In July this year, the Italian news service

ANSA reported that patients had claimed to have received the Stamina therapy in Georgia. It did not identify them or detail their clinical experience. But on 24 October, 52-year-old Andrea Zicchieri appeared in a tell-all report on the Italian television news channel La7, complete with video documentation. Zicchieri, who has a motor-neuron disease, said he had paid €18,000 (US\$20,000) for a cycle of three treatments in Tbilisi between July and September that had been arranged with Vannoni. Zicchieri told La7 that his brother had gone to Tbilisi in May to donate bone-marrow cells that were used in the treatment, and that Vannoni was present while the treatment was carried out. He said that the treatment had not worked.

Zicchieri, who has now been called to give evidence to the police, repeated the same details to *Nature*, providing screenshots of messages between himself and Vannoni, and adding that he had sought out Vannoni knowing that the treatment was controversial. “Even if there was only 1% chance of hope, I wanted to take that chance,” he said.

His account is consistent with talk on the social-media accounts of patient groups that support Vannoni’s work. According to photographs supplied by Zicchieri, the clinic was the Mardaleishvili Medical Center in Tbilisi, which also features on a patient-group site dedicated to the Stamina treatment. When *Nature* rang the centre for comment, a man said that Vannoni had been there and would return in January, and that all discussions about stem-cell therapy should go through Vannoni and not the clinic. He didn’t answer on further attempts to make contact.

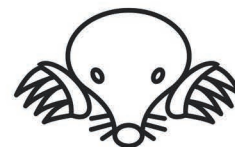
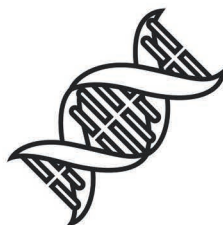
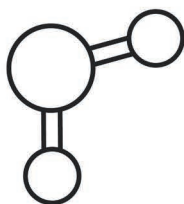
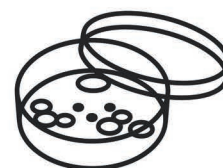
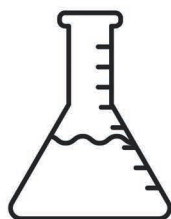
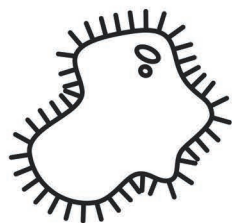
Postings on two of Vannoni’s Facebook pages — previously reported in Italian media — suggest that he is now working abroad. In February, in response to a question about his whereabouts, Vannoni posted that he was “abroad to continue Stamina”. On another page, he posted a message in June thanking people for their birthday wishes, and noting that “sometimes one must abandon a battle to win a war. In short — to change one’s country. #750”. Underneath the post, commentators said that the #750 means that Vannoni is now claiming to have made 750 stem-cell infusions — many more than he had made in Italy. (An earlier message, posted on 24 December 2015, ends “#450”).

Vannoni’s therapy involves modifying stem cells taken from the bone marrow of a patient or relative, and injecting them into the patient. In Italy, Vannoni has used this approach in people with a range of conditions, from Parkinson’s disease to muscular dystrophy. Recent cases of similar treatments have led to cancers, notes Elena Cattaneo, a neuroscientist at Italy’s University of Milan who was among those who worked to stop Vannoni and who is now an Italian senator. That Vannoni has apparently started again is “a disgrace”, she says. “Governments and health institutes should do more to inform patients about these sorts of therapies.” ■

MAURO UJETTO/NURPHOTO/ZUMA WIRE

I CAN HAZ MORE SCIENCE EMOJI?

Host of nerd icons proposed.



BY KATHERINE BOURZAC

Texting is set to reach new heights of scientific expression. A group of science enthusiasts and designers is proposing to add new science emoji to the official list of the icons. If approved by the organization

that oversees the list, text messages could soon feature DNA, a mole or a conical flask. Existing official science emoji include a telescope and a magnifying glass. Group members say that they want scientists to see their interests reflected in the official emoji offerings. ■

See go.nature.com/2g6ei5e for a longer version of this story.



Rwanda has improved the quality of its home-grown science.

and infant deaths, new HIV infections, AIDS deaths and mother-to-child HIV transmission (S. Nsanzimana *et al. BMC Med.* **13**, 216; 2015).

Kagame has also used his authority to ensure that science conducted by local and foreign researchers promotes domestic development. In 2012, the country's ministry of health published guidelines compelling all foreign science projects to strengthen Rwandan research capacity, for example by training its scientists or building infrastructure. Rwandan researchers routinely appear as first and last authors on studies conducted in the country, in contrast to other African nations where local researchers often don't benefit from foreign collaborations.

"You can't just be a global health researcher who drops in, gets some data, publishes it with your name as first author and never comes back," says epidemiologist Edward Mills of the Institute for Health Metrics and Evaluation in Seattle, Washington, who is an adjunct professor at the University of Rwanda in Kigali.

DAVID EVANS/GETTY

HARD LINE ON HEALTH

Staff at clinics and ministries who fail in their roles to meet stringent health targets can also be reassigned or sacked. In July, for instance, Kagame removed Agnès Binagwaho from her post as minister of health. The highly regarded paediatrician earned the US\$100,000 Roux Prize in April 2015 for using data to improve public health, but was let go after malaria cases in the country quadrupled to 2 million between 2012 and 2015.

Human-rights groups have chafed at Kagame's authoritarian tendencies, but he has kept corruption low in Rwanda compared with other sub-Saharan African nations. This made the country a favourite of donors for much of the 2000s. But that shifted after 2013, when organizations such as the US Institute of Medicine questioned donors' generosity towards Rwanda over nations with much higher HIV burdens.

In response, donors recalibrated. In 2014, the Global Fund to Fight AIDS, Tuberculosis and Malaria began using a new formula that allocated funding in part according to a country's burden of disease. In June, the fund said that it had reworked its formula again in response to protests from countries such as Rwanda that argued they were being punished for their success.

Despite the funding constraints and bureaucratic restrictions, Rwandan researchers who train abroad often return home. Kateera earned his medical degree in Uganda, where universities have partnerships with prestigious institutions in Europe, the United States and Asia. With one-third of the population and one-ninth the area of Uganda, Rwanda doesn't offer the same opportunities. But Kateera feels that he can make a difference in Rwanda: "You can make a big impact and measure it much more easily here compared to in a larger country," he says. ■

DEVELOPMENT

Donors slash aid to Rwanda

Progress on public health and science prompts a shift in funding to more-troubled nations.

BY ERIKA CHECK HAYDEN

Rwanda has made major public-health strides since the country's genocide against the Tutsi people ended in June 1994, but declines in foreign aid now threaten that progress.

Donors such as the US President's Emergency Plan for AIDS Relief and the Global Fund to Fight AIDS, Tuberculosis and Malaria have reduced assistance to Rwanda by 40% over the past three years. The situation will be a hot topic at the annual meeting of The World Academy of Sciences in Kigali on 14–17 November.

"If the decline in funding continues, there are a lot of things to lose rather than to gain," says Sabin Nsanzimana, who manages initiatives on HIV and other blood-borne diseases at the Rwanda Biomedical Center in Kigali, which runs the country's health programmes.

The declining foreign aid is part of two broader trends in development: redirecting money to countries that have the highest number of sick people, and urging developing countries to fund more of their own development work. The former has reduced aid to Rwanda, a small country that has slashed the

incidence of diseases such as HIV.

Like many other developing nations, Rwanda doesn't have the resources to move money from other priority areas such as education into health to fill the aid gap, says Nsanzimana. Even if sub-Saharan African nations were to more than triple their spending on HIV in the next five years, he notes, a recent study found that most could only raise half the money they need to end the epidemic's global threat by 2030 — a goal set by the Joint United Nations Programme on HIV/AIDS (M. Remme *et al. Soc. Sci. Med.* **169**, 66–76; 2016).

Fredrick Kateera, director of research for the Rwanda office of the non-profit organization Partners in Health, says that funding cuts could imperil the research that is needed to fight diseases such as malaria in Rwanda and elsewhere. "Setting up surveillance systems costs just as much money as just giving out drugs and bed nets," he says.

Rwanda has long been seen as a prime example of how science can aid development. After the genocide in 1994, President Paul Kagame invested in building roads, developing high-speed Internet access and applying science to local problems. The country slashed maternal

SCIENCE PUBLISHING

Reviews open up

Trials suggest growing support for transparent peer review.

BY EWEN CALLAWAY

When Kevin Sinclair reported that 13 cloned sheep his lab had studied lived long and healthy lives, he wanted to be as transparent as possible about what has been a controversial research area. Sinclair had invited journalists to see the sheep while the experiment was in progress at the University of Nottingham's School of Biosciences in Loughborough, UK. And when his paper was published this July, he agreed that its peer-review reports should also be made public (K. D. Sinclair *et al.* *Nature Commun.* 7, 12359; 2016).

The developmental biologist was taking part in a trial by *Nature Communications*, in which the journal offered authors the option to have their reviews published. The goal was to find out whether scientists would see the practice as a way to make research more reliable and egalitarian — or as a needless nod to transparency that could harm peer review.

A few journals, such as *PeerJ*, the *BMJ* and *F1000Research*, already embrace open peer review in different forms. Some forbid it. Other publishers and journals, including *Nature Communications*, are treating the practice as a frigid swimming pool: they are dipping their toes in the water, but are reluctant to plunge in.

So far, scientists seem willing to give open peer review a try. On 10 November, *Nature Communications* announced that around 60% of its authors in

2016 had agreed to have their reviews published, and that it would continue to offer scientists the option — although would not make it mandatory. (Reviewers can choose to withhold their names, but cannot otherwise influence the process, apart from declining to take part in an 'open review' paper.)

Meanwhile, an unpublished online survey funded by the European Commission (EC) has found that more than half of its 3,062 respondents thought that open peer review should become routine, although they expressed some qualms about specifics. One challenge is that open peer review means different things to different people, says Anthony Ross-Hellauer, an information scientist at the Göttingen State and University Library in Germany, who ran the survey for the EC-funded 'OpenAIRE' project on open science. Some think that it implies only naming the reviewers, but not

making their reports available, whereas others think that unsigned reports should be public. "It makes it really difficult to talk about what works in what circumstances if we're not using the same language," Ross-Hellauer says.

Some scientific communities seem to embrace open-review reports more readily than others, notes Joerg Heber, executive editor of *Nature Communications*. During his journal's trial, authors on more than 70% of eligible papers in ecology and evolution, molecular biology and Earth sciences adopted open reports, whereas physics papers saw the lowest uptake.

OPEN EXPERIMENTS SPREAD

For nearly two years, Dutch publisher Elsevier has published unsigned peer-review reports for five of its titles. On the basis of that trial, Elsevier plans to bring open review to other journals next year. And some journals allow reviewers to post their pre-publication reviews at other websites. One of those sites is Publons.com, which encourages scientists to make their peer reviews public, if journals permit.

Advocates say that the benefits of open peer review are straightforward. "It's about making the process fairer and more transparent, so people can be held accountable if something goes wrong," says Jonathan Tennant, a palaeontologist at Imperial College London and communications director at ScienceOpen, an open-research publishing network. But Stephen Heard, an ecologist at the University of New Brunswick in Fredericton, Canada, has misgivings. He worries about scientists posting peer-review reports without the knowledge or permission of reviewers. Heard also says that if he knew his reviews would end up public, it would make the job harder because he would feel obliged to cut back on technical language. "I would inevitably do fewer reviews," he says.

It's also not clear who actually reads the reports once they are made public, Heard says. Heber says that download figures will be assessed by *Nature Communications*. (The journal is published by *Nature's* publisher, Springer Nature; *Nature's* news and comment team is editorially independent of the publisher's research editorial teams.)

Sinclair has not yet heard from anyone who has read the review report for his paper. And although his lone experience was encouraging, he is not ready to be an evangelist for open peer review. "I think most people, and I would include myself, are apprehensive," Sinclair says. ■

"It's about making the process fairer and more transparent."



THE HOSTILE OCEAN THAT SLOWED CLIMATE CHANGE

The waters of the Southern Ocean have delayed global warming by absorbing much of the excess heat and carbon generated by humanity.

But that might not last.

BY JEFF TOLLEFSON

Joellen Russell wasn't prepared for the 10-metre waves that pounded her research vessel during an expedition south of New Zealand. "It felt like the ship would be crushed each time we rolled into a mountain of water," recalls Russell, an ocean modeller at the University of Arizona in Tucson. At one point, she was nearly carried overboard by a rogue wave.

But what really startled her was the stream of data from sensors analysing the seawater. As the ship pitched and groaned, she realized that the ocean surface was low in oxygen, high in carbon and extremely acidic — surprising signs that nutrient-rich water typically found in the deep sea had reached the surface. As it turned out, Russell was riding waves of ancient water that had not been exposed to the atmosphere for centuries.

Although controversial when she encountered it back in 1994, this powerful upwelling is now recognized as a hallmark of the Southern Ocean, a mysterious beast that swirls around Antarctica, driven by the world's strongest sustained winds. The Southern Ocean absorbs copious amounts

Strong winds help to pull up ancient water from the ocean bottom.

of carbon dioxide and heat from the atmosphere, which has slowed the rate of global warming. And its powerful currents drive much of the global ocean circulation.

The hostile conditions have kept oceanographers at bay for decades, but a new era of science is now under way. Researchers from around the world are converging on the region with floats, moorings, ships, gliders, satellites, computer models and even seals fitted with sensors. The goal is to plug enormous data gaps and bolster understanding of how the Southern Ocean — and the global climate — functions. Doing so could be key to improving predictions of how quickly the world will warm, how long the Antarctic ice sheet will survive and how fast sea levels will rise.

“It’s been amazing to see this explosion of information,” says Arnold Gordon, an oceanographer at Lamont-Doherty Earth Observatory in Palisades, New York, who led some of the early Southern Ocean surveys in the 1960s. “New technologies are allowing us access to these remote areas, and we are far less dependent on driving a ship through the sea ice.”

Already, initial data from an array of ocean floats suggest that upwelling waters could be limiting how much CO₂ the Southern Ocean absorbs each year. This raises new questions about how effective these waters will be as a brake on global warming in decades to come.

“The Southern Ocean is doing us a big climate favour at the moment, but it’s not necessarily the case that it will continue doing so in the future,” says Michael Meredith, an oceanographer with the British Antarctic Survey in Cambridge, UK. Meredith is heading a series of expeditions over the next five years to help document the uptake of heat and carbon. “It really is the key place for studying these things.”

TRACKING CARBON

The mysteries of the Southern Ocean have beckoned explorers for centuries, but the unique geography of the region makes it a perilous place for ships. There are no landmasses to tame the winds and waves that race around the planet at 60° S. And the ice surrounding Antarctica is notorious for engulfing wayward vessels, including Ernest Shackleton’s *Endurance* in 1915.

Scientists only started to realize how important the region is for controlling global climate in the 1980s, when several groups were trying to explain what had caused atmospheric CO₂ concentrations to drop by about one-third during the last ice age and then later rise. Oceanographer Jorge Sarmiento at Princeton University in New Jersey realized that changes in circulation and biology in the Southern Ocean could help to cool and warm the planet¹.

Three decades later, Sarmiento is leading an effort to gather the first real-time data on the chemical and biological processes that govern carbon in the Southern Ocean. The US\$21-million Southern Ocean Carbon and Climate Observations and Modeling Project (SOCCOM) has already deployed 51 of a planned 200 robotic floats that bob up and down in the upper 2,000 metres of the Southern Ocean. Building on the global Argo array, which consists of more than 3,700 floats collecting temperature and salinity data, the SOCCOM floats also measure oxygen, carbon and nutrients.

With the new data, Sarmiento and his team can test their models and refine estimates of how CO₂ moves between the seas and the sky. Indirect evidence suggests that the Southern Ocean is a net carbon sink and has absorbed as much as 15% of the carbon emissions emitted by humanity since the industrial revolution. But at some times of year and in specific places in this region, carbon-rich surface waters release CO₂ into the atmosphere.

Now, researchers are getting some of their first glimpses in near-real time of what happens in the Southern Ocean, particularly in winter. “Right off the bat, we are seeing CO₂ fluxes into the atmosphere that are much greater than we had estimated before,” Sarmiento says. “It’s just revolutionary.”

The unpublished analysis is based on just 13 floats that have been in the water for at least a year, so the question now is whether the higher CO₂ emissions during winter represent larger trends across the entire Southern Ocean.

“It’s pretty tantalizing,” says Alison Gray, a postdoctoral researcher at Princeton who is leading the study. “It would imply that potentially there is a much weaker carbon sink in the Southern Ocean than has been estimated.”

Hints of something similar have been seen before. In 2007, a team led by Corinne Le Quéré, now director of the Tyndall Centre for Climate Change Research in Norwich, UK, published a study in *Science*² indicating that the rate of carbon uptake by the Southern Ocean decreased between 1981 and 2004. The authors blamed the changes on the winds that encircle the Antarctic continent. The speed of those winds had increased during that time, probably as a result of the hole in the stratospheric ozone layer over Antarctica and possibly because of global warming. Stronger winds are better able to pull up deep, ancient water, which

releases CO₂ when it reaches the surface. That would have caused a net weakening of the carbon sink.

If that trend were to continue, atmospheric CO₂ levels would rise even faster in the future. However, a study in *Science*³ last year found that the carbon sink started to strengthen in the early 2000s (see ‘The unreliable sink’).

Le Quéré says it’s unclear whether that rise in CO₂ absorption is a return to normal or a deviation from the long-term weakening of the sink. Regardless, she says, it’s now clear that the Southern Ocean might be much more fickle than scientists thought.

SOCCOM floats will probably help researchers to answer these questions, but it could be years before they can say anything concrete about trends. Nor is Le Quéré convinced that the new network of floats will provide enough detail. In a paper published in July⁴, she found that models of carbon uptake by the Southern Ocean depend strongly on assumptions about the structure of the food web there. She says that climate scientists need to improve their understanding of the type and timing of phytoplankton and zooplankton blooms if they are going to get their climate projections right. “In my view, that’s the next frontier,” she says.

WARMING WATERS

Carbon is only part of the story in the Southern Ocean. Scientists are also beginning to pin down what happens to all the heat that gets absorbed there.

The Southern Ocean is the starting point for a network of currents that carry water, heat and nutrients throughout the ocean basins. Near Antarctica, surface waters normally grow cold and dense enough to sink to the bottom of the ocean, forming abyssal currents that hug the sea floor as they flow north into the Pacific, Atlantic and Indian oceans.

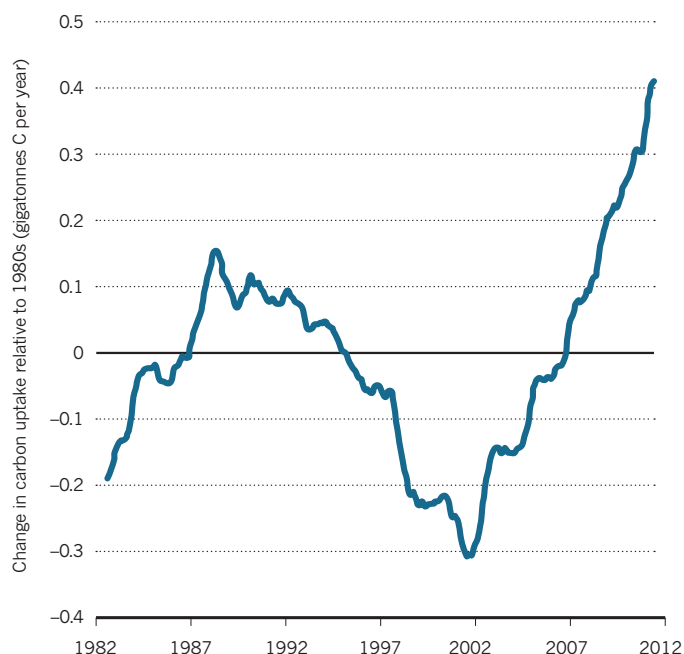
Much of what scientists know about these currents comes from ship surveys conducted every decade or so since the early 1990s. In 2010, when researchers analysed data from the surveys, they found a pronounced warming trend in abyssal waters, which were somehow absorbing about 10% of the excess heat arising from global warming⁵.

The level of warming in the deep ocean came as a surprise, and researchers have proposed several explanations that centre on the

“THE SOUTHERN OCEAN IS DOING US A BIG FAVOUR AT THE MOMENT, BUT IT’S NOT NECESSARILY THE CASE THAT IT WILL CONTINUE.”

THE UNRELIABLE SINK

The amount of carbon dioxide absorbed by the Southern Ocean has fluctuated markedly. In the 1990s, the region lost some of its capacity to take up the greenhouse gas. It started to suck in more in 2002, but this trend could reverse.



Southern Ocean. One factor could be that surface waters around Antarctica have become less salty, in part because of an increase in summer rainfall over the ocean. Fresher surface water is less dense, so that change would choke the supply of cold water sinking to the sea floor to feed the bottom currents. “The deep water warms up because it’s not getting as much cold-water replenishment,” says Gregory Johnson, an oceanographer with the National Oceanic and Atmospheric Administration (NOAA) in Seattle, Washington, who co-authored the 2010 analysis.

An as-yet-unpublished analysis, based on initial data from the third round of ship surveys, finds similar trends, but researchers have longed for more frequent measurements to provide a fuller picture. That could happen if a proposed international project moves forward. Called Deep Argo, this would be an array of floats that regularly dive all the way to the bottom of the ocean. Johnson is involved in a US consortium that is testing 13 floats in a basin off the coast of New Zealand, and another nine south of Australia.

Others are using moorings to monitor deep water flows. Since 1999, Gordon has maintained an array of moorings in the Weddell Sea, one of the main areas where cold surface waters sink to form ocean bottom currents. He has seen the deep water growing less salty in some areas, but the long-term trends are not clear⁶.

“We are really only scratching the surface of how bottom waters are changing, and how that is impacting the large-scale global ocean circulation,” he says.

ALONG THE EDGE

In January 2015, oceanographers aboard the Australian icebreaker *Aurora Australis* were cruising off the coast of Antarctica when they were presented with a unique opportunity. Following a crack in the sea ice, they were able to reach the edge of the Totten Glacier, one of the biggest drainage points for the East Antarctica ice sheet. No other expedition had reached within 50 kilometres of the glacier.

The team deployed floats and gliders into the waters around and underneath the glacier, which is 200 metres thick at its front edge. What they found came as a shock. The water at the front of the glacier was

3 °C warmer than the freezing point at the base of the glacier.

“We always thought Totten was too far away from warm water to be susceptible, but we found warm water all over the shelf there,” says Steve Rintoul, an oceanographer at the Antarctic Climate and Ecosystems Cooperative Research Centre in Hobart, Australia.

Scientists had already shown^{7,8} that warm-water currents are undercutting the West Antarctic ice sheet in many areas along the peninsula where the glaciers extend into the ocean. But Rintoul says that this expedition provided some of the first hard evidence that these same processes are affecting East Antarctica, raising new questions about the longevity of the mammoth ice sheets that blanket the continent.

There is no clear answer yet for what is driving the warming of these near-surface currents. Some explanations invoke changes in the winds over the Southern Ocean and the upwelling of warm waters. Others focus on fresher surface waters and an expansion of sea ice in some areas. The combination of extra sea ice and fresher surface waters could create a kind of cap on the ocean that funnels some of the warmer upwelling water towards the coast.

“Every scientist, including me, has their favourite explanation,” Gordon says. “But that’s how science works: the more you observe, the more complicated it gets.”

Finding the answers may require recruiting some of Antarctica’s permanent residents. Meredith’s team at the British Antarctic Survey plans to equip Weddell seals with sensors so that the animals can collect water measurements as they forage below the sea ice along the continental shelf. This zone has particular importance because it is precisely where cold water begins its descent into the abyss.

“The processes that happen in that shelf region are very important on a global scale, but measuring them is very difficult,” Meredith says. “The seals sort of transcend that barrier.”

The Weddell seals are just one component of the expedition’s arsenal. The team will also send autonomous gliders under the sea ice on pre-programmed routes to collect temperature and salinity data down to depths of 1,000 metres. Measurements taken from ships will help fill in the picture of what happens in this crucial region around Antarctica — and how it relates to the rest of the global ocean circulation.

Getting the data is only half the challenge. Ultimately, scientists need to improve their models of how currents transport heat, CO₂ and nutrients around the globe. Even armed with better measurements, results suggest that modellers have a way to go.

An analysis of data from the ship surveys suggests that upwelling ocean water does not rise in a simple pattern near Antarctica. Rather, it swirls around the continent one and a half times before reaching the surface. And Sarmiento’s team at Princeton found that only the highest-resolution models could accurately capture that behaviour. Sarmiento says that it could be a while before the models can simulate what really happens in this region, but he is confident that day will eventually arrive.

For Russell, it’s as if scientists are at last lifting the veil on the Southern Ocean. After she returned from her maiden voyage in 1994, she turned to modelling because there wasn’t enough data at the time to quantify the effects of the upwelling she encountered. Today she has it both ways. Russell is heading the modelling component of the SOCCOM project, and she is getting more data than she ever dreamt of.

“It’s just a wonderful time to be an oceanographer,” she says, “even as we are carrying out this really scary geophysical experiment on our planet.” ■

Jeff Tollefson writes for *Nature* from New York.

1. Sarmiento, J. L. & Toggweiler, J. R. *Nature* **308**, 621–624 (1984).
2. Le Quéré, C. et al. *Science* **316**, 1735–1738 (2007).
3. Landschützer, L. et al. *Science* **349**, 1221–1224 (2015).
4. Le Quéré, C. et al. *Biogeosciences* **13**, 4111–4133 (2016).
5. Purkey, S. G. & Johnson, G. C. *J. Clim.* **23**, 6336–6351 (2010).
6. Gordon, A. L., Huber, B., McKee, D. & Visbeck, M. *Nature Geosci.* **3**, 551–556 (2010).
7. Joughin, I., Smith, B. E. & Medley, B. *Science* **344**, 735–738 (2014).
8. Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. *Geophys. Res. Lett.* **41**, 3502–3509 (2014).

COMMENT



MEDICINE A glorious and gory history of New York's Bellevue hospital **p.354**

HISTORY The scientific currents in Lewis Carroll's words **p.356**

POLICY Scientists react to UK plan to reform research and universities **p.357**

OBITUARY Joseph L. Birman, human-rights activist and physicist, remembered **p.358**

REUTERS/ANDY CLARK



Morning fog masks the view as a landscaper rakes the grass on the Vancouver Convention Center's living roof in Vancouver, Canada.

Carbon is not the enemy

Design with the natural cycle in mind to ensure that carbon ends up in the right places, urges **William McDonough**.

Carbon has a bad name. The 2015 Paris climate agreement calls for a balance between carbon dioxide emissions to the atmosphere and to earth-bound carbon sinks¹. Climate Neutral Now, a United Nations initiative, encourages businesses and individuals to voluntarily measure, reduce and offset their greenhouse-gas emissions by 2050. The American Institute of Architects has challenged the architecture community worldwide to become carbon neutral by 2030. The Carbon Neutral Cities Alliance, an international network of urban-sustainability directors, aims to slash its cities' greenhouse-gas emissions by 80% by 2050.

'Low carbon', 'zero carbon', 'decarbonization', 'negative carbon', 'neutral carbon', even 'a war on carbon' — all are part of the discourse. If we can reduce our carbon emissions, and shrink our carbon footprint, the thinking goes, we can bring down the carbon enemy. It's no wonder that businesses, institutions and policymakers struggle to respond.

But carbon — the element — is not the enemy. Climate change is the result of breakdowns in the carbon cycle caused by us: it is a design failure. Anthropogenic greenhouse gases in the atmosphere make airborne carbon a material in the wrong place, at the wrong dose and for the wrong duration. It is

we who have made carbon toxic — like lead in our drinking water or nitrates in our rivers. In the right place, carbon is a resource and a tool.

Carbon dioxide is the currency of photosynthesis, a source of Earth's capacity for regeneration. Soil carbon is the guarantor of healthy ecosystems and food and water security. Carbon atoms are the building blocks of life. Wool, cotton and silk are carbon compounds, as are many industrial polymers and pure 'supercarbons' such as diamonds and graphene.

After 30 years of designing sustainable buildings and landscapes that manage carbon, I believe it is time to breathe new ►

► life into the carbon conversation. Rather than declare war on carbon emissions, we can work with carbon in all its forms. To enable a new relationship with carbon, I propose a new language — living, durable and fugitive — to define ways in which carbon can be used safely, productively and profitably. Aspirational and clear, it signals positive intentions, enjoining us to do more good rather than simply be less bad.

WORDS DRIVE ACTIONS

It is easy to lose one's way in the climate conversation. Few of the terms are clearly defined or understood. Take 'carbon neutral'. The European Union considers electricity generated by burning wood as carbon neutral — as if it releases no CO₂ at all. Their carbon neutrality relies problematically on the growth and replacement of forests that will demand decades to centuries of committed management². Another strategy is to offset fossil-fuel use by renewable-energy credits — this still means an increase in the global concentration of atmospheric CO₂.

Even more confusing is the term 'carbon negative'. This is sometimes used to refer to the removal of CO₂ from the atmosphere. For example, Bhutan's prime minister has indicated that his country is carbon negative, because its existing forests sequester more CO₂ than the country emits and Bhutan exports hydroelectric power (see go.nature.com/2es9lgt). But aren't trees having a positive effect on atmospheric carbon, and hydroelectric power a neutral one?

Carbon sequestration is a long-sought goal. It requires two elements: a way to capture carbon from the atmosphere or a chimney and a way to store it safely and permanently. But some so-called carbon-storage methods are paradoxical. For example, in enhanced oil recovery, CO₂ is injected

into rock formations to flush out remnant crude oil, which is eventually burned.

At the same time, enterprises are starting to announce their hopes to be 'carbon positive' by, for example, producing more renewable energy than their operations require, or by sequestering carbon through planting trees.

Such terms highlight a confusion about the qualities and value of CO₂. In the United States, the gas is classified as a commodity by the Bureau of Land Management, a pollutant by the Environmental Protection Agency and as a financial instrument by the Chicago Climate Exchange.

A new language of carbon recognizes the material and quality of carbon so that we can imagine and implement new ways forward (see 'The new language of carbon').

"To enable a new relationship with carbon, I propose a new language."

It identifies three categories of carbon — living, durable and fugitive — and a characteristic of a subset of the three, called working carbon. It also identifies

three strategies related to carbon management and climate change — carbon positive, carbon neutral and carbon negative.

START WITH THE SOIL

How do we work with the carbon cycle to preserve and enhance the benefits it naturally provides? From the soil up.

Carbon is at the heart of soil health. In healthy ecosystems, when plants convert CO₂ into carbon-based sugars — liquid carbon — some flows to shoots, leaves and flowers. The rest nourishes the soil food web, flowing from the roots of plants to communities of soil microbes. In exchange, the microbes share minerals and micronutrients

that are essential to plants' health. Drawn into the leaves of plants, micronutrients increase the rate of photosynthesis, driving new growth, which yields more liquid carbon for the microbes and more micronutrients for the fungi and the plants. Below ground, liquid carbon moves through the food web, where it is transformed into soil carbon — rich, stable and life-giving. This organic matter also gives soil a sponge-like structure, which improves its fertility and its ability to hold and filter water.

This is how a healthy carbon cycle supports life. This flow kept carbon in the right place in the right concentration, tempered the global climate, fuelled growth and nourished the evolution of human societies for 10,000 years.

Many soil researchers believe it could do so again. Ecologist and soil scientist Christine Jones, founder of the Amazing Carbon Project, describes the "photosynthetic bridge" between atmospheric carbon and liquid carbon, and the "microbial bridge" between plants and biologically active, carbon-rich soils as twin cornerstones of landscape health and climate restoration³.

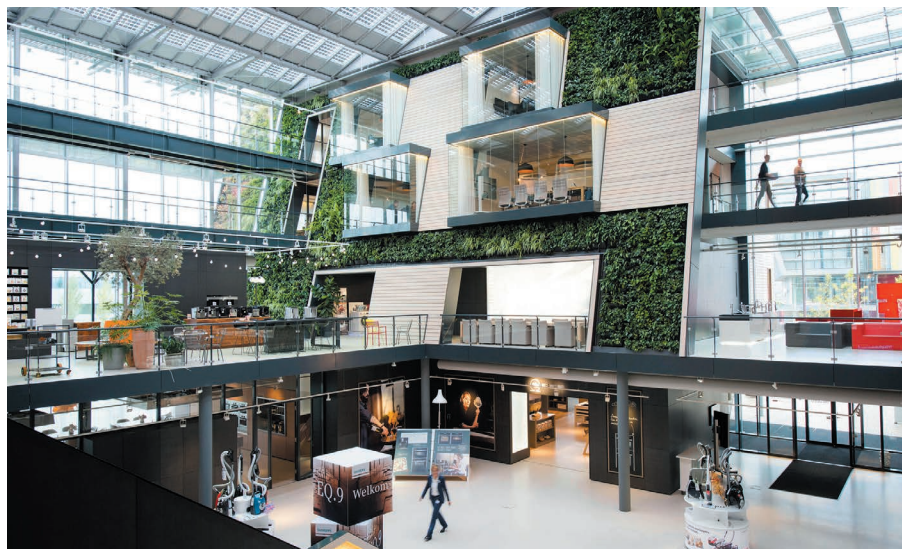
David Johnson at the New Mexico State University Institute for Energy and the Environment in Las Cruces has studied the carbon-microbial bridge⁴. He found that the most important factor for promoting plant growth and cultivating soil carbon was not added nitrogen or phosphorus but the carbon inputs from other plants.

DESIGN FOR LIVING

Let's keep those carbon bridges open on all landscapes — rural and urban. Let's use carbon from the atmosphere to fuel biological processes, build soil carbon and reverse climate change. Let's adopt regenerative farming and urban-design practices to increase photosynthetic capacity, enhance biological activity, build urban food systems, and cultivate closed loops of carbon nutrients. Let's turn sewage-treatment plants into fertilizer factories. Let's recognize carbon as an asset and the life-giving carbon cycle as a model for human designs.

All designs — from products to buildings, cities and farms — could be carbon positive. This may take a century, but that's how long it took us to get into our current carbon calamity. The sooner we start, the better. By 2030, our exuberantly urbanizing planet is expected to convert more habitat and farmland into cities than all previous urban growth combined. More than 2 billion urbanites will live in homes, attend schools and work in factories that are not yet built⁵. Despite these challenges, there are models of hope.

In 1989 my architecture firm designed a day-care facility in Frankfurt, Germany, based on 'a building like a tree' that could be operated by children, who would move



A four-storey atrium with indoor and outdoor living green walls helps to provide clean air to Park 20|20's Bosch Siemens Experience Centre in the Netherlands.

SANDER VAN DER TORREN FOTOGRAFIE

THE NEW LANGUAGE OF CARBON

Too much carbon in the atmosphere is damaging. Instead, it should be retained in durable forms such as plastic and wood or in living organisms. Recycling materials and nurturing the soil ensure that carbon ends up in the right places in the right amounts.

FUGITIVE CARBON

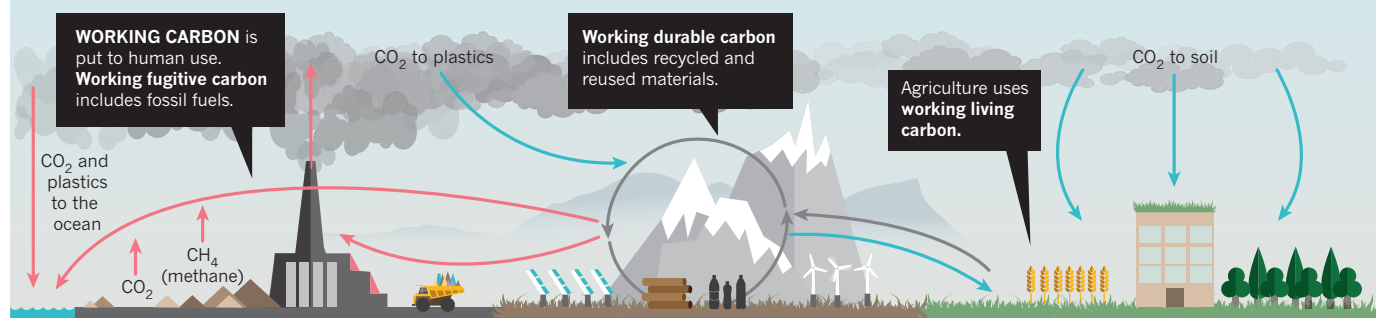
Has ended up somewhere unwanted and can be toxic. It includes carbon dioxide released into the atmosphere by burning fossil fuels, 'waste to energy' plants, methane leaks, deforestation, much industrial agriculture and urban development. Plastic in the ocean is fugitive carbon.

DURABLE CARBON

Locked in stable solids such as coal and limestone, or in recyclable polymers that are used and reused. It ranges from reusable fibre, such as paper and cloth, to building and infrastructure elements that can last for generations and then be reused.

LIVING CARBON

Organic, flowing in biological cycles, providing fresh food, healthy forests and fertile soil. It is something we want to cultivate and grow. Soil includes living carbon in the form of fungi, microbes, humus, legumes and grasses.



MANAGEMENT STRATEGIES

CARBON NEGATIVE

Actions that pollute the land, water and atmosphere with various forms of carbon. For example, releasing methane into the atmosphere or plastic waste into the ocean is carbon negative.

CARBON NEUTRAL

Actions that transform or maintain carbon in durable earthbound forms and cycles for use across generations; or renewable energy such as solar, wind and hydropower that do not release carbon.

CARBON POSITIVE

Actions that convert atmospheric carbon to forms that enhance soil nutrition or to durable forms such as polymers and solid aggregates. Also includes the recycling of carbon into soil nutrients from organic materials, food waste, compostable polymers and sewage.

solar shutters, open and close windows, grow food on roof terraces and irrigate the gardens with rainwater.

The idea of 'buildings like trees' and 'cities like forests' endured, and we started to approach our product, building and city designs as photosynthetic and biologically active, accruing solar energy, cycling nutrients, releasing oxygen, fixing nitrogen, purifying water, providing diverse habitats, building soil and changing with the seasons.

The Adam Joseph Lewis Center for Environmental Studies at Oberlin College in Ohio, which we designed, is a built example of this philosophy. It purifies its waste water and sewage in an on-site system that produces carbon-rich organic compost. This year the project is producing solar energy at an annual rate of 40% more than it needs. The building still relies on the electrical grid when solar energy is unavailable. Soon, with new and affordable on-site thermal and electric battery storage systems, buildings like this can be both carbon and energy positive.

In the Netherlands, Park 20|20 near Amsterdam applies these carbon-positive design strategies at the campus scale. Next door, the Valley at Schiphol Trade Park, the country's national hub for the circular economy, will scale these and many other innovations to create an urban ecology of work,

supply chains and collaborative spaces. The development will be a network of integrated buildings, landscapes and technical systems operating as a connected whole. Each building is oriented to the path of the Sun to maximize exposure during winter and shade during summer. Photovoltaic arrays and green roofs are the system's leaves and roots, harvesting renewable energy, absorbing and filtering water, producing food and providing habitat for other living things in a vibrant, sustainable business community.

The energy sector, too, can be generously carbon positive. SunPower, based in San Jose, California, and other solar providers are developing 'solar orchards' — power plants that perform as working farms. Rotating arrays of elevated solar panels shade the earth and provide habitat for grassland, which captures water, nitrogen and carbon to build soil health, can include legumes to fix nitrogen, and can provide food for grazing animals, in turn providing protein and wool. By design, the power plant generates an abundance of benefits: renewable energy, biodiversity, food, soil restoration, nutrient cycling, carbon sequestration, water conservation, fibre products, and agricultural and manufacturing jobs. Thus working durable carbon creates and supports living carbon while reducing fugitive carbon, all in an

economically robust and profitable model.

Such designs offer an inspiring model for climate action. It all starts with changing the way we talk about carbon. Our goal is simple and positive: a delightfully diverse, safe, healthy and just world — with clean air, soil, water and energy — economically, equitably, ecologically and elegantly enjoyed. ■

William McDonough is an architect and author specializing in sustainable development. He is founder of William McDonough + Partners and McDonough Innovation, and co-founder of MBDC (a design chemistry firm), in Charlottesville, Virginia, USA. He is also a founder of the Cradle to Cradle Products Innovation Institute.

e-mail: info@mcdonough.com

1. United Nations Adoption of the Paris Agreement: Annex (2015); available at <http://go.nature.com/2f7b9oue>
2. Upton, J. *Pulp Fiction: The European Accounting Error That's Warming the Planet* (Climate Central, 2015); available at go.nature.com/2f7b9oue
3. Jones, C. *WANTFA New Frontiers in Agriculture* September 100–104 (2013); available at go.nature.com/2f7b9oue
4. Johnson, D., Ellington, J. & Eaton, W. *PeerJ PrePrints* **3**, e789v1 (2015).
5. United Nations Department of Economic and Social Affairs *World Urbanization Prospects: 2014 Revision, Highlights* (United Nations, 2014); available at go.nature.com/2f7b9oue



BETTMANN/GETTY

Surgery at Bellevue hospital in New York City in the 1870s.

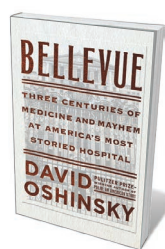
PUBLIC HEALTH

Gore and glory

David Dobbs extols a history of New York's Bellevue hospital, a crucible of discovery in medicine.

To many Americans, the name Bellevue signifies 'psychiatric facility' as much as Bedlam does to Britons. The psychiatric unit of the New York City public hospital gained fame from the stream of cultural icons passing through its portals. Writer Delmore Schwartz arrived in handcuffs after trying to strangle a hostile book reviewer; jazz great Charles Mingus checked in voluntarily, later composing the song *Lock 'Em Up* (*Hellview of Bellevue*). Yet, as historian David Oshinsky shows in his sweeping, eponymous chronicle, this oldest, busiest, most storied of New York hospitals deserves equal recognition as a fount of medical discovery.

From its beginnings as a colonial almshouse in 1736, Bellevue had important roles in one major advance after another. It opened the first US maternity ward (1799), nursing school (1873), children's clinic (1874), emergency department (1876), mental-health ward (1879) and, in the famed Carnegie Laboratory (1884), pathology lab. It was the first US medical centre to successfully tie off a femoral artery, and to report tuberculosis as treatable. It pioneered lasting innovations



Bellevue: Three Centuries of Medicine and Mayhem at America's Most Storied Hospital
DAVID OSHINSKY
Doubleday: 2016.

shunted from other New York hospitals. Many arrive shattered by infectious diseases, mental illness and poverty. This "municipal eyesore most New Yorkers couldn't imagine living without" has repeatedly had to ask the city to give yet more billions, which it always does. The challenges that routinely almost crush the institution

in medical education, public-health outreach and professional nursing; the use of ambulances, amputation, anaesthesia and antiseptics; and successful treatment of AIDS and Ebola. It has achieved all this while struggling to honour a pledge to treat all comers.

That pledge guarantees a huge patient load, heavy on "undocumented, uninsured, and undomiciled" people

also drive its boundless creativity.

Oshinsky draws deftly from this history to tell stories of an unlikely institution and Western medicine's halting, bloody march forward. He also reveals the torturous ambivalence over universal patient care in a nation where treatment decisions remain largely revenue-driven. To all of it he brings an eye for particulars, a knack for apt quotes and a talent for braiding multiple themes.

One such theme is how often medicine does harm because of compassionate and competitive pressures to do something. In the early nineteenth century, physicians were obliged to "cure quickly", as one complained, "or give place to a rival". Not even presidents were immune to the results of such pressure. In 1881, Bellevue surgeon Frank Hamilton inserted his dirty fingers into the wounds received by President James Garfield during an assassination attempt, probably contributing to Garfield's slow, gruesome death. Hamilton billed the nation US\$25,000 (almost \$600,000 today); he was paid \$5,000.

Oshinsky similarly enlivens Bellevue's more successful work in forensic pathology and AIDS treatment. A century ago, the hospital's Charles Norris, tall, gruff and confident, provided the archetype for the no-nonsense medical examiner. Spending family wealth to stock his labs and working obsessively to solve mysterious deaths, Norris played a prime part in making forensic pathology a medical and legal force in the United States. His best employee, chemist Alexander Gettler, became the mould for the geeky forensic ace of popular culture. Gettler discovered how to

tell whether a floating corpse had drowned or been tossed dead into the water, and how to use soil stains or thread fragments to tie a criminal to a crime. A fellow expert opined that Gettler “sent more criminals to the electric chair through his tests than any police detective”.

The AIDS chapter astonishes more grimly. New York, with a high incidence of intravenous drug use and promiscuous unprotected sex in the late 1970s, was hit hard by HIV. One of the first people to be recognized as having AIDS presented himself in 1980 to Fred Valentine, an infectious-disease specialist at Bellevue, with the rare pneumocystis pneumonia. A second case came days later. Both men soon died, as did eventually hundreds who followed them into Bellevue and nearby New York University Hospital. By 1985, one-fifth of Bellevue’s patients had HIV. At first, these hospitals, with others in San Francisco, California, and the Centers for Disease Control and Prevention in Atlanta, Georgia, fought a seemingly futile effort to contain this accelerating epidemic, even as cases began to be recognized in Africa and on other continents (T. Tansey *Nature* 533, 468; 2016).

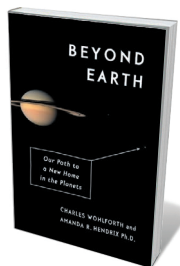
Oshinsky makes vivid the piteous condition of early patients, and caretakers’ struggles with fear of contagion. Bellevue staff’s sense of “therapeutic impotence”, he shows, finally gave way to exhausted relief in the mid-1990s. Teams led by Valentine and fellow medic David Ho separately developed similar antiretroviral (or ART) drug cocktails that inhibited the virus’s replication and made AIDS manageable.

Bellevue’s record shines bright amid the shameful behaviour of many in the US medical system around HIV and AIDS. In 1986, with more than 20,000 dead and tens of thousands more infected, the American Medical Association informed its members that it was legitimate to ignore these patients. The United States, Oshinsky notes, continues to ration care according to status and wealth — an ailment that the Obamacare medical-insurance statute salves but does not cure. This year, for instance, conservatives in the US Congress repeatedly sabotaged bills meant to fund the fight against the expanding Zika epidemic which, up to that point in the country, had affected mainly the US territory of Puerto Rico. The funding, requested by President Barack Obama’s administration in February, finally passed on 28 September.

Bellevue is rich in anecdote, history, personality and narrative. It is also an indictment of a society’s failure, almost 300 years on, to extend medical benefits to all, despite the efforts of this extraordinary hospital. ■

David Dobbs writes on science, medicine and culture. His books include *Reef Madness* and the short memoir *My Mother’s Lover*. e-mail: david.a.dobbs@gmail.com

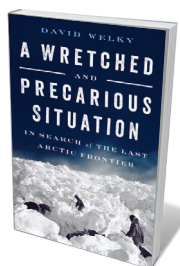
Books in brief



Beyond Earth: Our Path to a New Home in the Planets

Charles Wohlforth and Amanda R. Hendrix PANTHEON (2016)

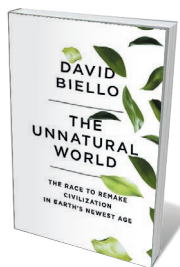
“They will go boating on lakes of liquid methane and fly like birds in the cold, dense atmosphere.” Life on Saturn’s moon Titan could prove exhilarating, suggest writer Charles Wohlforth and planetary scientist Amanda Hendrix in this synthesis of space-colonization science. Their thought experiment balances possible futures with a raft of facts on advances in spacecraft technology, robotics and space medicine. Crucially, they parse the push and pull between cautious governments and gung-ho entrepreneurs, concluding that the two may ultimately add up to a propulsive combination.



A Wretched and Precarious Situation

David Welky W. W. NORTON (2016)

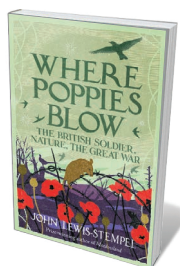
In 1906, polar explorer Robert Peary sighted a mysterious region northwest of Canada’s Arctic Archipelago. Dubbing it Crocker Land, he enlisted anthropologist Donald MacMillan and geologist Elmer Ekblaw as part of a US–Inuit team to explore the landmass under the aegis of New York’s American Museum of Natural History. What happened next, historian David Welky reveals in this engrossing account of the five-year effort (1913–17), involved not only the classic litany of illness, privation and howling blizzards, but a singularly bizarre finding about Peary’s original sighting.



The Unnatural World

David Biello SCRIBNER (2016)

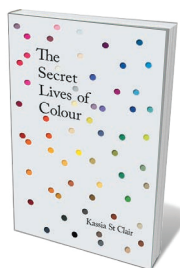
Forget the Anthropocene epoch, argues journalist David Biello. What we need is a broader “Anthropozoic” era, a lasting future anchored in inspired planetary stewardship and intelligent optimism. His lucid survey of researchers straining to contain today’s global environmental shifts (from climate change to dwindling marine biodiversity) mixes their personal scientific journeys with contextualizing discoveries. Among them are marine biologist Victor Smetacek, experimenting with iron fertilization to draw carbon dioxide into the oceans, and ecologist Erle Ellis, monitoring forestation with drones.



Where Poppies Blow: The British Soldier, Nature, the Great War

John Lewis-Stempel W&N (2016)

For traumatized, trench-bound British soldiers caught up in the carnage of the First World War, birdwatching and botany offered solace. So reveals John Lewis-Stempel in this riveting study drawing on verse, letters and field notes by men who served, from zoologist Dene Fry to poet Edward Thomas. He shows how observing the nesting larks that twittered above ‘no man’s land’ and natural cycles such as the seasons gave a sense of renewal, and how animal ‘troops’ were inspirationally loyal. A remarkable picture of a human bloodbath that took place amid phenomenally rich biodiversity.



The Secret Lives of Colour

Kassia St Clair JOHN MURRAY (2016)

Heliotrope, gamboge, umber: the names of colours are as luscious as the hues themselves. Kassia St Clair serves up a chromatic buffet of the chemistry, history and cultural associations of 75 dyes, pigments and shades, including encapsulations of optics and colour theory. The gorgeous, malodorous Turkey red, for instance, was made in a “tortuous process involving rancid castor oil, ox blood and dung”, and the Roman emperor Nero used a large emerald as “proto-sunglasses” while watching gladiatorial combat. [Barbara Kiser](#)

HISTORY

Untangling Alice

Gillian Beer reveals the currents in Lewis Carroll's worlds.

Lewis Carroll's lucid dreams draw in endless fresh contexts for interpretation as they continue to delight and disturb. How did Charles Dodgson — Carroll's real name — do it? The story used to be that this 'mediocre' English mathematician, isolated at Christ Church College, Oxford, somehow miraculously produced *Alice's Adventures in Wonderland* (1865) and *Through the Looking-Glass, and What Alice Found There* (1871). The unlikely tale persisted, perhaps because so little evidence seemed to have survived of Dodgson's reading; at his death, most of his books were hastily sold off.

But recent years have seen the publication of Edward Wakeling's *The Diaries of Lewis Carroll* (The Lewis Carroll Society, 1993–2007) and Charlie Lovett's *Lewis Carroll Among His Books* (McFarland, 2005) — a descriptive catalogue of the volumes that Dodgson owned or is known to have read. These have given me convincing external evidence of the range of his curiosity: across evolution and logic, dream theory and botany, mathematics and dictionaries, animal welfare and new ideas about language. I have also explored his fascination with parody and the satirical weekly magazine *Punch*. In my new book, *Alice in Space* (University of Chicago Press, 2016), I investigate the array of debates to which Carroll playfully responded, and so find lost jokes and poignant vanished references.

When I started writing it more than a decade ago, I wondered how far intuition and familiarity with Victorian intellectual culture should take me in asserting Carroll's participation in the ideas thronging around him. I had to rely on the Alice books for evidence of allusion and parodies. Now I have a fuller picture of how Carroll used fantasy to pursue thoughts — on radical mathematics and Boolean logic, for example — that he constrained in his professional life as a devout Euclidean (F. F. Abeles *Nature* 527, 302–304; 2015).

The Victorian culture within which the Alice books were written is largely invisible to us now. It was a period of immense intellectual upheaval in fields from mathematics to language theory, evolution and education. Carroll slips these ideas into

the layers of his jokes, sliding infant puns above learned references. He had a teasing openness to the ideas being pursued by his contemporaries in science, such as Thomas Henry Huxley, Charles Darwin, by logician John Stuart Mill and by lawmakers, photographers, museum-makers and novelists including George Eliot and Emily Brontë.

The Alice books move like quicksilver — sleek globules of sense splitting and straying through the maze. This was a period when possible worlds, probability and ideas of space as curved or flat were being eagerly discussed. What mathematician James Joseph Sylvester described as the “rumpling of the page” of 3D space fascinated, among others, physicist Hermann von Helmholtz, who observed that Euclid's geometry holds good for the plane, but not for “surfaces flexible without change of dimensions”. Thus Alice goes through the looking-glass.

Carroll's taste for games and play was shared by many of his contemporaries, and understood as essential intellectual stimulus. For instance, Sylvester's 1869 address to the British Association for the Advancement of Science emphasized the need to quicken the mind of students “with the doctrine of the imaginary and the inconceivable”. The logician Augustus de Morgan wrote in 1859 “All that is thinkable is possible; all that is impossible is unthinkable: that is, so far as our knowledge can go.” Carroll, who knew both men, put it this way in an encounter between Alice and the White Queen in *Looking-Glass*:

Alice laughed. “There's no use trying,” she said: “one can't believe impossible things.”

“I daresay you haven't had much practice,” said the Queen. “When I was your age, I always did it for half-an-hour a day. Why, sometimes I've believed as many as six impossible things before breakfast.”

The idea of rigorous training in believing the impossible nicely tilts at Victorian learning rituals in the style of Gradgrind, Charles Dickens's rigid pedagogue in *Hard Times* (1854). Nonsense was also valued by the physicist James Clerk Maxwell, who appreciated both the Alice books. His 1873 poem ‘Molecular evolution’ declared:

One of Alice's transformations.

What combination of ideas,
Nonsense alone can wisely form!
What sage has half the power that she has,
To take the towers of truth by storm?

Carroll put several systems in motion at once in his scenes, to produce new absurdities and insights. For example, the names of the snooty talking flowers in *Looking-Glass* parody the then-popular 1855 poem ‘Come into the garden, Maud’, and the fashionable ‘language of flowers’ for lovers. They also nod to a recent theory about the origin of language in animal sounds that Carroll's friend, the philologist Max Müller, scorned as ‘the bow-wow theory’:

“But what could it do, if any danger came?”

Alice asked.

“It could bark,” said the Rose.

“It says ‘Bough-wough!’” cried a Daisy.

“That's why its branches are called boughs!”

Alice, eager and adamant, is also a transformative figure. In *Wonderland* she can, after trial and error, manage the scale of her body: wish fulfilment for any child subject to the awkward demands of growing up. And the often improbable relation of childhood appearance to the later adult was being newly dramatized now that they could be set side by side in photographs. Transformation was a perturbing idea for many Victorians, with its threat of kinship between species (in the books, a baby turns into a pig; bottles, plates and forks become birds). All taxonomic systems seemed unstable in the wake of evolutionary theory — to say nothing of the threat to human exceptionalism. After Alice's neck grows, she encounters a pigeon defending its eggs. Hearing that little girls eat eggs, it announces, “then they're a kind of serpent”.

Carroll emphasizes in his sideways stories that everything has a voice and every thing a right to speak, be it queen, cat, door, Time, Mock Turtle, egg or child. The Unicorn in *Looking-Glass* remarks of Alice:

“I always thought they were fabulous monsters!” said the Unicorn. “Is it alive?”

“It can talk,” said Haigha solemnly.

The Unicorn looked dreamily at Alice, and said: “Talk, child.”

Talk is the stuff of life in these books: deft and provoking talk that draws on an array of then-current ideas and yet holds steady the surface tension of the story. ■

Gillian Beer is emeritus professor at the University of Cambridge, UK, and an honorary fellow of Clare Hall and Girton Colleges, Cambridge, and St Anne's College, Oxford. Alice in Space is published this month. e-mail: gpb1000@cam.ac.uk



Correspondence

UK research reform: protest now

The grass-roots campaign group Science is Vital (<http://scienceisvital.org.uk>) shares the Royal Society president's concerns over aspects of the government's Higher Education and Research Bill (see V. Ramakrishnan *Nature* **538**, 459; 2016).

The bill does not yet offer sufficient protection for the operational autonomy of the UK research councils. Neither does it provide legal guarantees that future reforms will take into account the views of Parliament or the research community.

We are also alarmed by mechanisms the bill could use to aid entry and exit of institutions from the higher-education 'market'. For example, the new Office for Students will have the power to revoke university status without parliamentary assent. Such powers undermine support for the autonomy of seats of learning and enquiry that have proved their cultural worth for generations.

In our view, verbal assurances from the Minister for Universities and Science are insufficient. The stated "primacy of scientific and academic decision-making" must be enshrined in the bill.

As the bill is readied for its third reading before being debated in the House of Lords, we call on the UK research community to contact their MPs urgently to relay concerns about the dangers in this proposed legislation (see go.nature.com/2fsmclc).

Stephen Curry *Imperial College London, UK.*

Jenny Rohn *University College London, UK.*

Andrew Steele *Francis Crick Institute, London, UK.*
s.curry@imperial.ac.uk

UK research reform: poor timing

Science and universities in the United Kingdom surely do not need a major and controversial

restructuring during the approach to Brexit (see *Nature* **538**, 5; 2016). The government's proposed Higher Education and Research Bill stands to erode university autonomy, downgrade individual research councils and concentrate executive authority over science into a single 'supremo'.

It is plainly desirable for the research councils to collaborate more smoothly. And ministers need better advice on apportioning funding between councils and on such matters as balancing small-scale, 'responsive mode' grants against large strategic initiatives. However, these inadequacies can be remedied without the wholesale reorganization envisaged in the bill. From my perspective as Astronomer Royal and former Royal Society president, the research councils work better than most government agencies and need only fine-tuning.

A good start would be to ensure that there is a senior independent voice in Whitehall by reviving the post of Director-General of Research Councils, supported by a strong advisory board.

Martin Rees *Institute of Astronomy, University of Cambridge, UK.*
mjr@ast.cam.ac.uk

UK research reform: get the facts straight

For universities with a royal charter, the UK government's proposed Higher Education and Research Bill does not "rip up an 800-year-old settlement" (*Nature* **538**, 5; 2016). And on a factual point, the University of Cambridge does not have a royal charter. Pope John XXII gave us formal recognition in 1318, and our privileges were confirmed by Parliament in 1571 through the Oxford and Cambridge Act.

A royal charter recognizes an institution or group of

individuals as a single legal entity, each with different responsibilities and rights. It is an exaggeration and simplification of the bill's proposals to say that it will revoke these. Instead, the bill would legally recognize institutional autonomy, the principle of dual support and the sector's diversity.

Evidence-based debate will improve the proposed legislation, for example through the Green Paper consultation and the Public Bill Committee (see also go.nature.com/2ejk7km). Discussions will continue as part of standard parliamentary procedure. The bill does not need to be thrown out to protect academic freedom.

Leszek K. Borysiewicz *University of Cambridge, UK.*
v-c@admin.ac.uk

Genetics boosts US–Cuban links

As economic doors open between the United States and Cuba, human genetics offers one promising area for scientific collaboration (see *Nature* **537**, 600–603; 2016). Reversing 50 years of restriction remains a formidable task — particularly with scant financial and human resources.

Community genetics is incorporated into Cuba's health-care system. As in the United States, prenatal genetic testing, screening of newborns and clinical genetics services are all available. Comparative studies on US and Cuban populations could help to clarify the genetic contribution to disease — for example, by revealing rare inherited genomic variants.

Miami is one of the best-positioned US cities to lead such scientific partnerships, given its social and cultural ties with Cuba (Florida is home to 66% of the US Cuban population). At the University of Miami's Leonard M. Miller School of Medicine, we are working to develop exchange programmes with medical

institutions in Cuba. These mutual learning opportunities should foster multidisciplinary research partnerships in the next generation of medical geneticists.

Rosario Isasi, Stephan L. Zuchner *University of Miami, Florida, USA.*

Roberto Cañete Villafranca *University of Medical Sciences, Matanzas, Cuba.*
risasi@miami.edu

Precision oncology is not an illusion

In our view, it is unreasonable to condemn personalized medicine for oncology on the basis of the limited success of a few trials (see V. Prasad *Nature* **537**, S63; 2016). We suspect that those failures were more likely to be caused by shortcomings in methodology.

With more than 40 precision-oncology drugs on the market, such therapies are helping tens of thousands of US patients by targeting specific molecular abnormalities. For example, mutations in the gene that encodes the epidermal growth-factor receptor are likely to occur in 10% of the 186,240 or so new cases of non-small-cell lung cancer predicted for 2016 (see go.nature.com/2fpxits). And the 8,220 people with chronic myeloid leukaemia (CML) predicted for this year will almost all carry the 'Philadelphia' chromosomal translocation (see go.nature.com/2fbbarj).

There are precision-oncology drugs to counteract both defects. Indeed, the life expectancy of people with CML is now starting to approach that of the general population (S. Saussele *et al.* *Blood* **126**, 42–49; 2015).

Edward Abrahams *Personalized Medicine Coalition, Washington DC, USA.*

Stephen L. Eck* *Astellas, Northbrook, Illinois, USA.*
eabrahams@personalizedmedicinecoalition.org
**Competing interests declared; see go.nature.com/2efcnj.*

Joseph L. Birman

(1927–2016)

Physicist who helped hundreds of scientists to escape oppression.

I first met Joseph L. Birman in 1979. He attended an unofficial seminar in Moscow for scientists like me who had lost access to academic institutions in the Soviet Union because of our political views or because we had applied to leave the country. We gathered in the apartment of a computer scientist who was under KGB surveillance (and later spent five years in prison and exile for anti-Soviet actions). This was before the Internet and social media. National borders were closed; we felt isolated and threatened by our government.

Although few Western academics dared attend our gatherings, there were many more people than chairs. Birman, a tall man, sat uncomfortably on the floor, trying to find space for his legs. Then, as always, he was talkative and cheerful: everybody's uncle.

Birman, who died on 1 October, was born in New York City on 21 May 1927, the grandson of Jewish immigrants from Russia. In 1943, he graduated from the Bronx High School of Science, famously an incubator of prominent researchers. He received a bachelor's in science from the City College of New York and a doctorate in theoretical physics from Columbia University in 1952, going on to work on the optical properties of semiconductors at GTE Laboratories in New York. A decade later, he became a professor at New York University, and in 1974, he joined the faculty of City College, where he remained until his death.

Sharply experimental in his thinking and prohibitively mathematical, Birman demonstrated how the branch of mathematics known as group theory can be applied to understand transitions between crystal phases and to predict light scattering and other optical properties of solids.

He leveraged the respect he gained from seminal papers in the 1960s and 1970s into advocacy for hundreds of scientists. In a letter endorsing Birman for the Andrei Sakharov Prize of the American Physical Society (APS), which recognizes scientists who fight for human rights, Iranian physicist Hadi Hadizadeh wrote: "His efforts to get me released from detention and solitary confinement in 2001 will not be forgotten by me, my family, and many scientists worldwide." Winning the award in 2010, Birman was delighted to see his name attached to that of the notable Soviet dissident and nuclear physicist.

Birman's trips to the Soviet Union began



in the 1970s with official invitations from the Soviet Academy of Sciences. During those trips he learned about the plight of Jewish scientists in the country. They were often denied promotion, travel abroad and positions at top research institutions. Applying for an exit visa frequently resulted in loss of employment, but rarely in permission to leave. Open protests led to arrests and imprisonment.

Birman used his travels and eminence to challenge the heads of Soviet research institutions on behalf of scientists caught in this plight. It is thanks to the efforts of him and his colleagues that I did not end up in jail, despite multiple KGB interrogations, and was finally allowed to leave the Soviet Union.

In the early 1990s, when many scientists in Russia were finally allowed to emigrate, Birman helped to establish the Program for Refugee Scientists in the United States, raising funds from private foundations. This supported visiting positions for more than a hundred émigré scientists in US universities and gave them time to secure permanent positions in industry and academia.

Birman played a crucial part, along with particle physicist Robert Marshak, in recovering a generation of Chinese physicists lost to Mao Zedong's cultural revolution in the 1960s and 1970s. During this time, most scientific research

ceased, concepts such as Einstein's theory of relativity were denounced as bourgeois and scientists were sent to do manual labour in the countryside. In 1983, Birman and Marshak travelled to Beijing on behalf of the APS and signed an agreement with the Chinese Academy of Sciences and the Ministry of Education that brought more than 60 middle-aged physicists to work in labs throughout the United States for up to three years. Many leaders of Chinese physics are alumni of that programme, and the scientific cooperation between Chinese and US physicists that now exists evolved largely from it.

When the programme came to an end in the tragic aftermath of the Tiananmen Square protests in 1989, Birman redirected his efforts to achieve justice for Chinese scholars who openly spoke their minds. He would get phone calls and even surprise visits from Chinese scientists. He welcomed and did everything in his power to help these people, counselling them on how to manage their careers and providing contacts and recommendations.

Birman chaired human-rights committees at the APS and the New York Academy of Sciences and in that capacity wrote hundreds of letters to heads of governments, kings and religious leaders. He publicized cases of unjustly imprisoned scientists. For more than 40 years, he served as vice-chair of the Committee of Concerned Scientists dedicated to protecting human rights and scientific freedom around the world.

Joe met Joan Sylvia Lyttle when they were both at graduate school. They married in 1950 and had three children. She became a professor of mathematics at Columbia University. The day before Joe died, he and Joan had spent hours discussing a potential overlap between her work on topology and his model of how the phase of a particle is influenced by its trajectory through space.

As much as Joe loved physics, getting a scientist out of prison had infinitely greater value to him than any scientific achievement. As a physicist and a humanitarian, Joe's influence touched so many lives. He will be dearly missed by his friends and remembered by hundreds of people he helped. ■

Eugene M. Chudnovsky is a distinguished professor of physics at the City University of New York and co-chair of the Committee of Concerned Scientists.
e-mail: eugene.chudnovsky@lehman.cuny.edu

DAN Z. JOHNSON

OPTICAL PHYSICS

Clear directions for random lasers

Random lasers use disordered structures to produce light, which is usually emitted in many directions. A random laser that can produce a collimated beam offers a wide range of applications, from imaging to security scanning.

DIEDERIK S. WIERSMA

A special class of laser called a random laser is cheap, easy to fabricate and has emission properties that, to a certain extent, resemble those of a regular laser¹. However, light from a random laser is usually emitted in an uncontrolled way, over a broad range of angles (Fig. 1a)². Writing in *Optica*, Schönhuber *et al.*³ report a breakthrough in this respect — namely, the realization of a random laser whose directionality of emission can be controlled. The authors even create a collimated beam, while at the same time maintaining the rich spectral behaviour that characterizes random lasers.

The essential difference between regular and random lasers is the way in which the ‘optical modes’ are formed. In a regular laser, light is typically reflected back and forth between two parallel mirrors, one of which is partially transparent so that some of the light can get through. If the distance between these mirrors matches the wavelength of the light (or a multiple of it), a resonance will occur — just like a violin string that emits sound at a wavelength that matches its length. Optical modes correspond to the wavelengths or, equivalently, frequencies of light at which these resonances exist. We now know that even the most complex disordered photonic structures exhibit such resonances, which allows their optical response to be described in terms of modes⁴. A wealth of these disordered structures is currently used for light generation by random lasers.

For a pair of parallel mirrors, the spatial distribution of light at the resonance frequencies is simple: it consists of a series of maxima and minima along a straight line. By contrast, the corresponding light distributions for disordered photonic structures are complex and can differ in nature, ranging from being highly spatially localized to extending over the entire structure of interest⁵. Depending on the type of structure, the number of optical modes can vary from a few to very many, each having a characteristic light-intensity spectrum. Under normal conditions⁶, the emission of a random laser is the result of a delicate interplay of all of its modes and their corresponding spectra.

Schönhuber and colleagues managed to

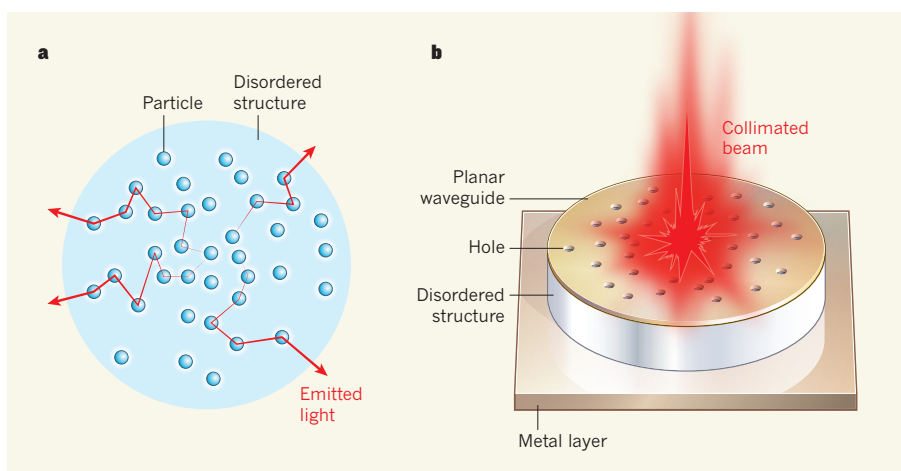


Figure 1 | A collimated beam from a random laser. **a**, In a random laser, light is amplified as it is transferred between particles in a disordered photonic structure. The light from a random laser is usually emitted in an uncontrolled way, over a broad range of angles. **b**, Schönhuber *et al.*³ sandwiched the photonic structure between a layer of metal and a planar waveguide, which restricted the propagation of light to a plane. The authors then inserted several small holes in the waveguide at random positions. By carefully optimizing the design of the waveguide and the properties of the light, the authors produced a collimated laser beam.

produce collimated light from a random laser by making cunning use of the laser’s optical modes. The authors used a planar waveguide — a structure that restricts the propagation of light to a plane — in which several small holes had been inserted in random positions (Fig. 1b). This type of configuration has two main advantages over previous studies^{4–6}. First, the in-plane scattering of light from the holes produces disordered modes whose spectra have varying spatial distributions. Second, the holes allow a tiny fraction of the light to scatter out of the plane and be emitted from the laser.

By carefully adjusting parameters such as the density and distribution of holes in the waveguide, Schönhuber and collaborators achieved the optimal in-plane confinement of light, while ensuring a sufficient amount of emission out of the plane. The same type of configuration had previously been proposed for enhancing the efficiency of thin-film solar cells⁷, which is perhaps not surprising because good absorbers of light can also be used as efficient emitters.

Next, Schönhuber *et al.* adjusted the polarization and phase of the laser light to obtain a

beam that was perpendicular to the waveguide and almost perfectly collimated. This concept had already been explored theoretically⁸ and has now been put into practice. At the same time, the authors maintained the advantages that come with random photonics: ease of fabrication, miniaturization and emission of photons that are coherent (in-phase) and that have a wide range of wavelengths.

Previous studies that produced light from a random laser using a disordered array of holes in a semiconductor membrane were limited to emissions in the mid-infrared range⁹ (frequencies of a few tens of terahertz; 1 THz is 10^{12} Hz). But Schönhuber and colleagues’ random laser can produce light at longer wavelengths (frequencies of a few THz), which is ideal for applications in areas as diverse as biomedical imaging, environmental monitoring and security scanning. The authors’ device could also be used in manufacturing because common packaging materials such as cardboard and plastics are transparent to terahertz radiation¹⁰.

However, one should bear in mind that using random optical modes also means that the emission spectra will be random.

Fortunately, the spectra produced by the authors are stable over time, and reproducible because there is limited competition between the different modes. But the fine structures of these spectra are disordered and contain a random collection of spikes in intensity. Using a slightly different waveguide produces a spectrum that has a distinct fine structure. Therefore, truly broadband emission, which includes light of all wavelengths within a given bandwidth, could be obtained only by putting together a large collection of such random lasers.

The intrinsic disordered nature of random laser emission could, however, be turned into an advantage. A large array of miniature emitters could serve as a light source with an intricate spectrum, the intensity spikes of which could be switched on and off. There is also an opportunity for applications in the field of security, in which the unique fine structure of the individual spectra could be used for tagging objects. One example is vehicle recognition (both military and civilian). The authors'

collimated beam, combined with its detailed and unique spectrum, could be used to identify vehicles at large distances that would otherwise be invisible to the human eye and undetectable using electronic equipment. The complex emission spectra and detailed fine structure would also have advantages over regular, single-frequency laser light in distinguishing weak signals from background noise — for instance, in imaging applications.

Random lasers are becoming more ordered and regular lasers are benefiting from disorder. Where is the field going? Random lasers have turned out to be less mysterious than initially thought, and we have gained a much better understanding of their rich and complex behaviour. The enormous range of materials that have been shown to exhibit random laser emission could suggest that random lasers will one day be more common than their regular counterparts. The clever use of disordered optical modes will help us to take advantage of the fascinating properties of random lasers

and tame their behaviour to the extent that is needed for practical applications. ■

Diederik S. Wiersma is at the *European Laboratory for Non-linear Spectroscopy (LENS), University of Florence, 50019 Sesto Fiorentino, Italy, and the National Institute for Metrological Research, Turin, Italy.*
e-mail: wiersma@lens.unifi.it

1. Cao, H., Ling, Y., Xu, J. Y., Cao, C. Q. & Kumar, P. *Phys. Rev. Lett.* **86**, 4524–4527 (2001).
2. Wiersma, D. S. *Nature Photon.* **7**, 188–196 (2013).
3. Schönhuber, S. *et al. Optica* **3**, 1035–1038 (2016).
4. Wang, J. & Genack, A. Z. *Nature* **471**, 345–348 (2011).
5. Fallert, J. *et al. Nature Photon.* **3**, 279–282 (2009).
6. Leonetti, M. & López, C. *Appl. Phys. Lett.* **102**, 071105 (2013).
7. Vynck, K., Burrelli, M., Riboli, F. & Wiersma, D. S. *Nature Mater.* **11**, 1017–1022 (2012).
8. Hirsch, T., Lierz, M., Pogany, D., Mintert, F. & Rotter, S. *Phys. Rev. Lett.* **111**, 023902 (2013).
9. Liang, H. K. *et al. Adv. Mater.* **25**, 6859–6863 (2013).
10. Tonouchi, M. *Nature Photon.* **1**, 97–105 (2007).

This article was published online on 9 November 2016.

MOLECULAR BIOLOGY

Mature proteins braced by a chaperone

Hsp70 chaperone molecules help other proteins to fold, and were thought to bind mainly to unfolded proteins. Single-molecule experiments now suggest that Hsp70s can also stabilize almost fully folded proteins. [SEE LETTER P.448](#)

QINGLIAN LIU & ELIZABETH A. CRAIG

Cell function depends not only on the proper folding and localization of proteins when they are synthesized, but also on maintaining the well-being of fully folded (mature) proteins, and on their disposal when they are irretrievably damaged¹. Seventy-kilodalton-sized heat-shock proteins (Hsp70s) are a major class of ubiquitous molecular 'chaperones' that have key roles in maintaining protein homeostasis^{2,3}. Hsp70s interact with short stretches of amino-acid residues in extended conformations, such as those in unfolded proteins⁴, and thus prevent such proteins from misfolding and aggregating. But on page 448, Mashaghi *et al.*⁵ report that Hsp70s are more flexible in their binding ability than was previously appreciated: they also interact with intermediates in folding processes, and with nearly fully folded proteins. This suggests that Hsp70s play a bigger part in protein homeostasis than was thought.

Analysing how protein folding is facilitated by molecular chaperones has been challenging — particularly for those chaperones, such as Hsp70, for which nucleotide binding and

hydrolysis drive cycles of interactions with proteins. Such chaperone–protein interactions are transient, and the conformations adopted by the substrate proteins are often highly variable and change over time. This is particularly true for the intermediates formed during protein folding.

Single-molecule analytical techniques can overcome these challenges, and thus are powerful approaches for addressing chaperone function. For example, the force generated by optical tweezers — a laser device that can hold and move microscopic objects — is ideal for pulling apart the folded structures of mature proteins to produce fully or partially unfolded proteins. Subsequent relaxation of the pulling force enables the refolding process to be observed. Chaperones, co-chaperones (proteins that assist chaperones in their functions) and nucleotides can be added to the molecule being studied to probe their effect on protein refolding and stability.

Mashaghi *et al.* used optical tweezers to analyse how DnaK, an Hsp70 from the bacterium *Escherichia coli*, interacts with maltose-binding protein. Surprisingly, they found that DnaK binds to and stabilizes nearly

fully folded and partially folded proteins, in addition to unfolded proteins. These results suggest that Hsp70s not only assist the folding of unfolded and misfolded proteins, but also help to maintain the conformation of mature proteins — thus maintaining their function. This preventive maintenance for 'healthy' mature proteins has not previously been suggested as a role for Hsp70 chaperones.

Although examples of Hsp70 binding to mature proteins have been reported before, the chaperone was thought to interact with extended loops in proteins⁴. Such thinking is consistent with structural studies^{4,6} that revealed Hsp70's peptide-binding cleft to be so constrained that only an extended, unstructured polypeptide segment of a protein can be accommodated in it. The substrate-binding domain (SBD) of Hsp70 is dedicated to binding polypeptide substrates, and comprises about half of the chaperone molecule. A subdomain of the SBD, known as SBD β , contains the peptide-binding cleft, and another subdomain, SBD α , forms a lid over the bound peptide segment (Fig. 1). SBD α thus stabilizes the chaperone–substrate interaction, but without forming any substantial direct contact with the substrate.

This model of substrate binding has provided a satisfying mechanistic explanation for how Hsp70 aids protein folding, especially the early steps in which nascent proteins consist of unfolded chains. Mashaghi *et al.* now find that the SBD α lid is particularly important in stabilizing almost-mature protein substrates — an intriguing observation that raises the question of how Hsp70s bind to such substrates to increase their stability. A structure of an Hsp70 in complex with a mature substrate would help to provide insight into the mechanics of this.

The substrate-binding activity of Hsp70s

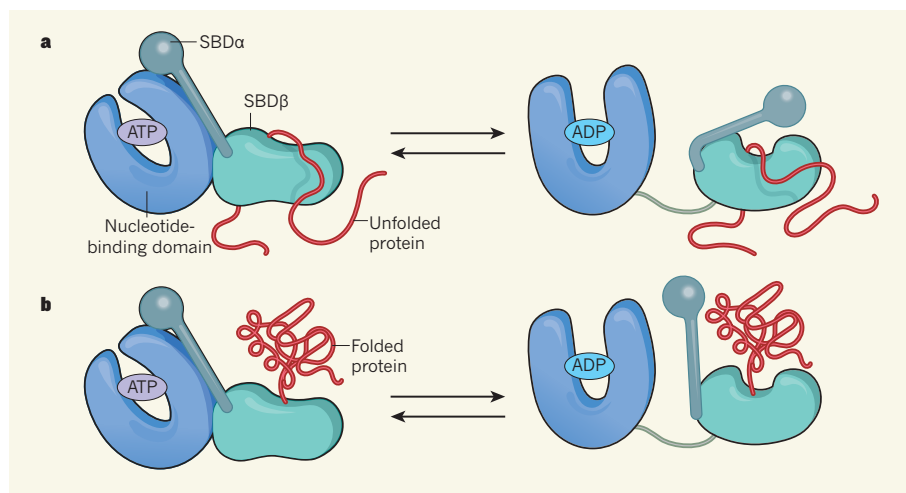


Figure 1 | Different roles of Hsp70s in protein folding. **a**, Seventy-kilodalton heat-shock proteins (Hsp70s) are molecular ‘chaperones’ that were thought to mainly assist the folding of unfolded, nascent protein chains. Unfolded proteins can bind to a cleft in a region of Hsp70 known as substrate-binding domain-β (SBDβ), and another region (SBDα) forms a lid that closes over the unfolded protein chain. The nucleotide-binding domain binds to ATP, which can be hydrolysed to form ADP. The lid is predominantly open when ATP is bound and closed when ADP is bound. **b**, Mashaghi *et al.*⁵ report single-molecule experiments with an Hsp70, and find that it interacts with an almost fully folded protein. This suggests that Hsp70s have roles in maintaining the structures of almost fully folded proteins.

depends on which nucleotide is bound^{7–11}: either ATP or the product of ATP hydrolysis, ADP. When ATP is bound, the peptide-binding pocket is predominantly open, allowing easy, rapid access for substrates. When ADP is bound, the peptide-binding pocket is predominantly closed — reducing substrate access but enhancing the stability of the chaperone–substrate interaction after the substrate is bound. The ATP-bound state therefore probably binds large, mostly folded substrates much more readily than the ADP-bound state. Both ATP binding and hydrolysis are crucial to the conformational changes of Hsp70 that allow it to bind and release substrates as they fold.

It is therefore surprising that Mashaghi and colleagues find that Hsp70 is more effective in stabilizing the (nearly) mature conformation of maltose-binding protein when ADP is the predominant nucleotide present, rather than ATP. However, the authors note that the ratio of ADP to ATP in cells is high during heat stress, and therefore propose that the observed stabilization is beneficial to the cell under such conditions — it might provide a way of maintaining proteins in their functional, folded conformations for little expenditure of energy. This explanation makes sense, but more work is needed to understand the physiological relevance of the results.

Mashaghi and co-workers’ study provides a wealth of information, but the limitations of the analysis should be kept in mind. The authors tested a single Hsp70: DnaK, a well-studied model of these chaperones. But even though Hsp70s are highly evolutionarily conserved, there are substantial differences in both their amino-acid sequences and their specific cellular functions. Does the observed

stabilization of folded protein substrates also occur in the Hsp70s of eukaryotic organisms (which include plants, animals and fungi)?

Moreover, Hsp70 activity in protein folding depends on the cooperation of two types of co-chaperone: J-proteins and nucleotide-exchange factors (NEFs)^{12,13}. These co-chaperones, especially NEFs, have much more diverse functions and amino-acid sequences than Hsp70s, so how do they fit into this expanded view of Hsp70 function? And most of Mashaghi and colleagues’ single-molecule assays involved just one kind of protein substrate. How generalizable are the

observed effects for other substrates?

Finally, it must be remembered that the unfolding force generated by optical tweezers is artificial. It is not clear how representative these findings are of the situation in cells. More experiments, from single-molecule studies to *in vivo* investigations, will therefore be needed to better understand the diverse roles of Hsp70s. But in the meantime, Mashaghi *et al.* have provided much food for thought. ■

Qinglian Liu is in the Department of Physiology and Biophysics, Virginia Commonwealth University, Richmond, Virginia 23298-0551, USA. **Elizabeth A. Craig** is in the Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706-1544, USA. e-mails: qinglian.liu@vcuhealth.org; ecraig@wisc.edu

- Balchin, D., Hayer-Hartl, M. & Hartl, F. U. *Science* **353**, aac4354 (2016).
- Mayer, M. P. & Bukau, B. *Cell. Mol. Life Sci.* **62**, 670–684 (2005).
- Boorstein, W. R., Ziegelhoffer, T. & Craig, E. A. *J. Mol. Evol.* **38**, 1–17 (1994).
- Clerico, E. M., Tilitsky, J. M., Meng, W. & Gierasch, L. M. *J. Mol. Biol.* **427**, 1575–1588 (2015).
- Mashaghi, A. *et al. Nature* **539**, 448–451 (2016).
- Zhu, X. *et al. Science* **272**, 1606–1614 (1996).
- Flynn, G. C., Chappell, T. G. & Rothman, J. E. *Science* **245**, 385–390 (1989).
- Schmid, D., Baici, A., Gehring, H. & Christen, P. *Science* **263**, 971–973 (1994).
- Qi, R. *et al. Nature Struct. Mol. Biol.* **20**, 900–907 (2013).
- Kityk, R., Kopp, J., Sinning, I. & Mayer, M. P. *Mol. Cell* **48**, 863–874 (2012).
- Zhuravleva, A., Clerico, E. M. & Gierasch, L. M. *Cell* **151**, 1296–1307 (2012).
- Kampinga, H. H. & Craig, E. A. *Nature Rev. Mol. Cell Biol.* **11**, 579–592 (2010).
- Bracher, A. & Verghese, J. *Front. Mol. Biosci.* **2**, 10 (2015).

This article was published online on 26 October 2016.

GEOCHEMISTRY

Ions surprise in Earth’s deep fluids

Models indicate that there are strong gradients in element concentrations and in the pH of fluids at the slab–mantle interface — a major discontinuity deep within Earth. This transforms our view of global geochemical transport. SEE LETTER P.420

DAVID DOLEJŠ

The chemical and isotopic composition of Earth’s rocks and minerals carries a fingerprint of how and where they formed. This information provides a unique record of the spatial and time scales of past planetary processes and the fluxes of the chemical elements involved. Interpreting this record

requires knowledge of the chemical conditions generated when aqueous fluids and minerals interact at extreme depths, but such knowledge has been hard to acquire. On page 420, Galvez *et al.*¹ report a method for computing the composition and chemical speciation of aqueous fluids in equilibrium with different rock types, and a predictive model for mass transport across the top of Earth’s subduction

zones — the regions at which portions of a tectonic plate are being forced underneath another plate.

Mass transport on our planet is driven by the motions of plates in Earth's outer shell (the lithosphere), and rocks formed at Earth's surface can be recycled to depths of up to 200 kilometres. As temperature and pressure rise with depth, the rocks undergo complex mineral transformations, including the release of volatile constituents such as water and carbon dioxide. These fluids are highly mobile in rock environments because of their low density and high buoyancy (Fig. 1), and act as transport agents for diverse chemical elements.

When the fluids penetrate major discontinuities in the Earth, such as the boundary between subducting slabs of lithosphere and the upper mantle, they act as a medium for complex dissolution–precipitation reactions, which allow subducted minerals to attain chemical equilibrium in their new rock environment. These reactions strongly alter element and isotope concentrations in the fluid and surrounding rocks, and thus chemically differentiate Earth's materials, dictate element mobility and define global element fluxes.

Understanding the reactivity of aqueous and carbon-containing fluids with minerals and rocks in Earth's deep interior has been a long-standing challenge for laboratory research and dynamic simulations. However, experimental advances^{2,3} in the past few years have allowed mineral solubilities and the stabilities of chemical species at extreme temperatures and pressures to be measured with unprecedented accuracy. In parallel, atomistic simulations^{4,5} have delivered new calibrations

of the electrostatic properties of water at such temperatures and pressures. These emerging data are essential for constructing sets of thermodynamic data for aqueous chemical species, which are used in simulations to predict fluid–mineral reactions and element transfer in natural systems. Such predictive simulations typically involve more than six chemical components, multiple mineral phases and tens of aqueous species in the fluid^{6,7}.

Nevertheless, obtaining robust information about the simultaneous exchange of multiple elements between minerals and fluids at Earth's extreme depths has not been possible — until now. Galvez and colleagues' computational

The authors realized that equilibria between mineral components and ionic species can be expressed using simple hydrolysis reactions.

method is based on the idea that the chemical potentials of major components of Earth's mineral assemblages (such as silica, alumina and sodium oxide) are the same regardless of whether they exist as aqueous species or as solids in rocks. Chemical potentials can be determined for all components within stable mineral assemblages, and dictate the chemical activity of each species and overall element concentrations in coexisting fluids at any pressure and temperature. The authors realized that equilibria between mineral components (oxides) and ionic species can be expressed using simple, independent hydrolysis reactions, and used this as the basis of their computational approach.

Galvez and co-workers' simulations reveal that, during subduction of Earth's oceanic crust, minerals containing hydroxyl or carbonate groups break down as the pressure and temperature increase, forming volatile-free solid phases and a free fluid. The fluid progressively dissolves the mineral phases, and thus acquires diverse cations, which make it moderately alkaline. Minerals in peridotite rock in the upper mantle surrounding the subducting slab also initially produce fluid that is alkaline, underpinned by the release of calcium cations from the mineral assemblage to the fluid. But when the temperature exceeds about 500 °C, the calcium cations become compatible with high-temperature minerals and therefore partition into the minerals from the fluid. The acidity of the aqueous fluid in the mantle thus increases, reaching weakly alkaline or near-neutral conditions. This phenomenon does not occur in the rocks of the subducting slab, and so a sharp acidity gradient is generated at the slab–mantle boundary.

But how does the difference in acidity affect fluid reactivity and chemical modification across this major geological interface? Galvez and colleagues' simulations show that infiltration of slab-derived fluid into the overlying mantle causes highly nonlinear effects: a rapid increase in fluid pH, followed by a sharp decrease. The infiltrated rocks respond in a highly nonlinear way, and may reach the limit of their ability to buffer or moderate the pH changes after coming into contact with a certain amount of fluid. This finding transforms our view of when the chemical composition of Earth's major geological domains is modified and where these processes are localized.

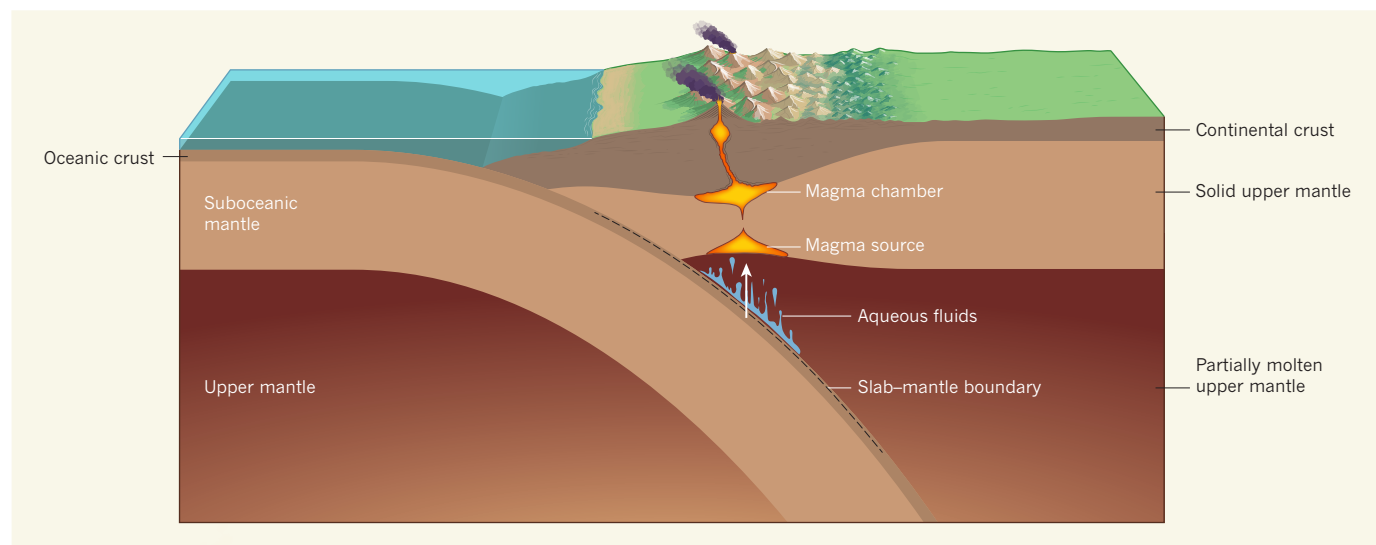


Figure 1 | Fluid movement at subduction zones. Subducting slabs of oceanic lithospheric plates (which consist of oceanic crust and suboceanic mantle) are exposed to increasingly high temperatures and pressures. Minerals in the oceanic crust that bear volatile constituents, such as water or carbonates, break down, and mobile fluids are released. The buoyant fluids rise from the slab and encounter the slab–mantle boundary — the chemical interface between rocks of the oceanic crust and the upper mantle. They

then rise and cool until they become incorporated into magma. The cycling of global elements depends on the solubility of chemical components in aqueous fluid in the slab, the reactivity of the multi-component fluid at the slab–mantle interface, and the incorporation into magmas of fluids with their load of dissolved material. Galvez *et al.*¹ report thermodynamic simulations that provide insight into the geochemical processes occurring in the deep slab and mantle.

The authors extended their predictive approach to fluids of even greater complexity that contain appreciable concentrations of ligand species, such as carbonates or chlorine. This allowed the chemical cycles of volatiles (including CO₂) and feedback relationships between ligand behaviour and metal mobility to be addressed.

The principal sources of carbonates at subduction zones are oceanic sediments and variably altered oceanic crust, whereas chlorine is derived from ocean water in the rock's pore and fracture space. At temperatures of between 200 and 700 °C, these ligands gradually form negatively charged species in fluids, and so concentrations of cations — including hydrogen ions, H⁺, which determine pH — in the fluid rise to balance the negative charge. Aqueous fluids associated with any slab rock type will therefore probably be weakly alkaline to almost neutral. Many dissolution–precipitation reactions are acid–base exchange processes (they involve the exchange of metal ions by H⁺), and so this relative increase in fluid acidity will probably promote the solubility and mobility of metals.

Earth's redox state changes between the upper continental crust and mantle by more than eight orders of magnitude^{8,9}, and exerts additional controls on the distribution of species in aqueous fluids. For example, iron(III) species are much less stable and are less soluble in water than their iron(II) counterparts, and CO₂ is more acidic than methane (CH₄). The effects of these characteristics act in conjunction to lower the net charge in the deep fluid and promote an increase in acidity. This perturbation propagates through different chemical species, resulting in new shifts of chemical equilibria and feedback relationships. Incorporation of redox equilibria into Galvez and colleagues' model will therefore probably allow previously unknown chemical gradients or interfaces to be identified in Earth's dynamic subduction systems. ■

David Dolejš is at the Institute of Earth and Environmental Sciences, University of Freiburg, 79104 Freiburg, Germany.
e-mail: david.dolejs@minpet.uni-freiburg.de

- Galvez, M. E., Connolly, J. A. D. & Manning, C. E. *Nature* **539**, 420–424 (2016).
- Wilke, M. *et al.* *Earth Planet. Sci. Lett.* **349–350**, 15–25 (2012).
- Bernini, D., Wiedenbeck, M., Dolejš, D. & Keppler, H. *Contrib. Mineral. Petrol.* **165**, 117–128 (2013).
- Pan, D., Spanu, L., Harrison, B., Sverjensky, D. A. & Galli, G. *Proc. Natl Acad. Sci. USA* **110**, 6646–6650 (2013).
- Sverjensky, D. A., Harrison, B. & Azzolini, D. *Geochim. Cosmochim. Acta* **129**, 125–145 (2014).
- Dolejš, D. & Wagner, T. *Geochim. Cosmochim. Acta* **72**, 526–553 (2008).
- Galvez, M. E., Manning, C. E., Connolly, J. A. D. & Rumble, D. *Earth Planet. Sci. Lett.* **430**, 486–498 (2015).
- Frost, D. J. & McCammon, C. A. *Annu. Rev. Earth Planet. Sci.* **36**, 389–420 (2008).
- Evans, K. A., Elburg, M. A. & Kamenetsky, V. S. *Geology* **40**, 783–786 (2012).

NEUROSCIENCE

Sleepy and dreamless mutant mice

Sleep in mammals consists of non-rapid-eye-movement and rapid-eye-movement sleep. A large genetic screen reveals that these two sleep states are altered in mice by mutations dubbed *Sleepy* and *Dreamless*. SEE ARTICLE P.378

DERK-JAN DIJK
& RAPHAËLE WINSKY-SOMMERER

Some people enjoy getting as much sleep as possible, whereas others claim to need only a few hours each night. Some people remember their dreams well, others don't. What is the genetic basis for these individual differences? In 'forward' genetic screens, genes are mutated at random and the effects of each mutation on a behaviour or biological trait of interest are assessed. On page 378, Funato *et al.*¹ report a large forward-genetic screen that identifies two genetic mutations that mediate variations in sleep patterns in mice.

Forward genetics is considered a gold-standard technique for identifying and characterizing previously unknown roles for genes. But one key prerequisite is that the trait of interest can be accurately assessed (phenotyped). More than 20 years ago², the phenotyping of circadian rhythms in mice through long-term recording of rest–activity cycles allowed researchers to use forward genetics to identify genes involved in these rhythms. But it has proved difficult to successfully apply forward genetics to the more complex process of mammalian sleep.

Sleep is much more than simply rest. It consists of two states — non-rapid-eye-movement sleep (NREM sleep; also known as deep sleep) and rapid-eye-movement sleep (REM sleep; also known as paradoxical sleep and sometimes referred to as dream sleep). Accurate phenotyping of sleep is challenging because it requires the identification and quantification of each of these sleep states through long-term, large-scale recordings of electrophysiological signals that measure brain waves and muscle activity.

Funato and colleagues rose to this challenge, recording patterns of sleep for two consecutive days in more than 8,000 mice carrying genetic mutations that were randomly induced in one of their parents. One group of mice slept, on average, for about 15 hours a day — more than 3.5 hours longer than the average of all screened mice, and much longer than previously documented for laboratory mice³. The authors, aptly, dubbed these mutant mice *Sleepy* (Fig. 1).

The researchers showed that NREM sleep is

specifically altered in *Sleepy* mice. The mutants had deeper baseline NREM sleep and reacted to sleep loss with more-intense NREM sleep than control mice. The mice responded to wake-promoting compounds such as caffeine, and Funato *et al.* found no evidence of behavioural abnormalities, ruling out defects in wake-promoting neural circuits as the cause of increased sleep. There were also no obvious abnormalities in the period of the animals' circadian rhythms. This finding is interesting; alterations in genes that control circadian rhythms have previously been shown⁴ to affect sleep, but the current experiment demonstrates that sleep- and circadian-based effects can be separated.

Funato and colleagues found that the mutation in *Sleepy* mice affects the *Sik3* gene, which has not previously been implicated in sleep. A single-nucleotide mutation prevents the inclusion of the 13th protein-coding region (exon) of *Sik3* in the SIK3 protein, presumably altering its activity. SIK3 is a protein-kinase enzyme that is expressed throughout the body and brain, and not just in known sleep-regulating areas — drawing a parallel with the clock genes that control circadian rhythms, which are expressed not only in the circadian master regulator in the brain, but also throughout the body. A strength of the paper is that the authors confirmed that *Sleepy* is indeed caused by skipping exon 13 of *Sik3*. By either introducing the same point mutation or using a refined genome-editing technology in mice, they recapitulated the *Sleepy* phenotype. Furthermore, they demonstrate that mutations in *Sik3* also affect sleep-like behaviours in fruit flies and roundworms, suggesting an evolutionarily conserved function for this gene.

From their screen, Funato *et al.* identified another group of mice, dubbed *Dreamless*, in which REM sleep time was reduced by an impressive 44% and occurred in shorter bouts than normal. It might be an overstatement to refer to these mice as *Dreamless* — some REM sleep remained, and humans also dream during NREM sleep⁵ (and some might doubt that mice experience dreams in the first place). Nonetheless, *Dreamless* mice have very little REM sleep compared with control mice³. The researchers demonstrated that the mutation in *Dreamless* mice affects the *Nalcn* gene, which

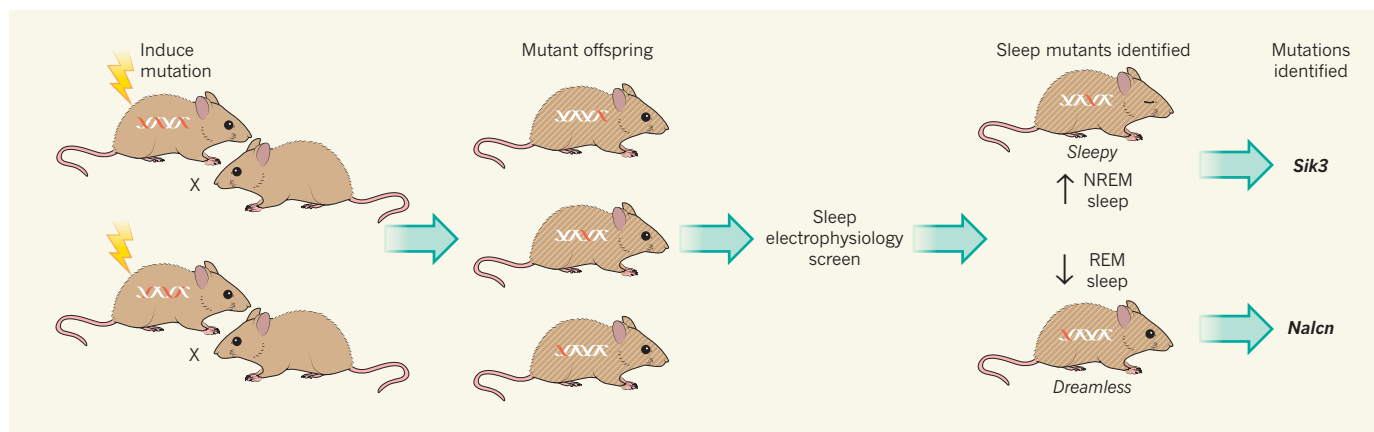


Figure 1 | Screening sleep. Funato *et al.*¹ performed a ‘forward’ genetic screen to search for mutations in genes that modulate sleep behaviours in mice. In forward genetics, subtle mutations (red) are randomly introduced into the DNA of mice, the mutant mice are crossed with wild-type partners, and the offspring produced are analysed for changes in a behaviour of interest. In this case, the authors analysed sleep in more than 8,000 offspring by taking

electrophysiological measurements of brain waves and muscle activity, and identified two groups of mice, dubbed *Sleepy* and *Dreamless*, that showed increased non-rapid-eye-movement (NREM) sleep and decreased rapid-eye-movement (REM) sleep, respectively. Subsequent genome sequencing revealed that the mutation in *Sleepy* mice affects a gene called *Sik3*; the mutation in *Dreamless* mice affects the gene *Nalcn*.

codes for an ion-channel protein and has previously been implicated in circadian control of neuronal excitability in fruit flies⁶.

In contrast to *Sleepy*, *Dreamless* mice have an altered circadian phenotype — the difference in activity levels between periods of rest and activity is reduced when these mice are subjected to constant darkness. This is of interest because REM sleep, but not NREM sleep, is under strong circadian control⁷. Finally, electrophysiologically assessed brain waves were altered in both mutants compared with controls, during REM and NREM sleep and during wakefulness. This finding, which implies that the mutations affect all vigilance states, could have been obtained only by the comprehensive phenotyping conducted.

What are the implications of this study? Sleep has multiple functions for the brain and the body, including regulation of learning, memory and metabolism⁸. Funato and colleagues’ identification of these new mutants opens up avenues for identifying the mechanisms by which sleep exerts these functions. For instance, whether the *Sleepy* and *Dreamless* mutations cause alterations in these traits can be addressed immediately.

The study might also have clinical implications. One obvious question is whether mutations in *SIK3* in people are implicated in a condition called idiopathic hypersomnia, in which people experience excessive sleepiness and sleep for long periods⁹. Another is whether mutations in *NALCN* have a role in increased age-related changes in cognitive decline, which in humans are predicted by reduced REM sleep¹⁰.

It is notable that Funato *et al.* identified no mice that did not sleep. This is in contrast to circadian rhythmicity, which can be abolished by a single mutation. Perhaps too many genes are involved in sleep for a forward-genetic screen to produce a sleepless mammal. Or

maybe sleepless mice are not viable, meaning that these mutations cannot be identified.

Overall, Funato and colleagues’ impressive study pays tribute to the importance and complexity of sleep by combining comprehensive phenotyping with state-of-the-art genetic and molecular tools. This strategy has unveiled new mutations that change the time spent in the two sleep stages that all mammals experience. ■

Derk-Jan Dijk and Raphaëlle Winsky-Sommerer are in the Surrey Sleep Research Centre, University of Surrey, Guildford, Surrey GU2 7XP, UK. e-mails: d.j.dijk@surrey.ac.uk; r.winsky-sommerer@surrey.ac.uk

1. Funato, H. *et al.* *Nature* **539**, 378–383 (2016).
2. Vitaterna, M. H. *et al.* *Science* **264**, 719–725 (1994).
3. Franken, P., Malafosse, A. & Tafti, M. *Sleep* **22**, 155–169 (1999).
4. Franken, P. & Dijk, D. J. *Eur. J. Neurosci.* **29**, 1820–1829 (2009).
5. Baylor, G. W. & Cavallero, C. *Sleep* **24**, 165–170 (2001).
6. Flourakis, M. *et al.* *Cell* **162**, 836–848 (2015).
7. Dijk, D. J. & Czeisler, C. A. *J. Neurosci.* **15**, 3526–3538 (1995).
8. Krueger, J. M., Frank, M. G., Wisor, J. P. & Roy, S. *Sleep Med. Rev.* **28**, 46–54 (2016).
9. Billiard, M. & Sonka, K. *Sleep Med. Rev.* **29**, 23–33 (2016).
10. Scullin, M. K. & Bliwise, D. L. *Sleep* **38**, 335–336 (2015).

This article was published online on 2 November 2016.

MATERIALS SCIENCE

Semiconductors that stretch and heal

Polymeric semiconductors have been prepared whose molecular properties make them stretchable and healable — a milestone in the development of sophisticated organic electronic surfaces that mimic human skin. SEE LETTER P.411

SIEGFRIED BAUER & MARTIN KALTENBRUNNER

Living organisms are made largely from soft substances that have the ability to stretch, and to heal when damaged. By contrast, everyday electronic appliances are made mostly of hard, brittle materials incapable of self-repair. Scientists desire to narrow this gap between the animate and digital worlds; this will require new approaches

in materials science. On page 411, Oh *et al.*¹ report their use of non-covalent-bonding mechanisms to make high-performance organic semiconductors that are intrinsically stretchable and healable. This allowed the authors to fabricate transistors — elementary electronic components used in amplifiers and logic circuits — entirely from stretchable materials in thin films, potentially opening the way to new generations of wearable electronic devices.

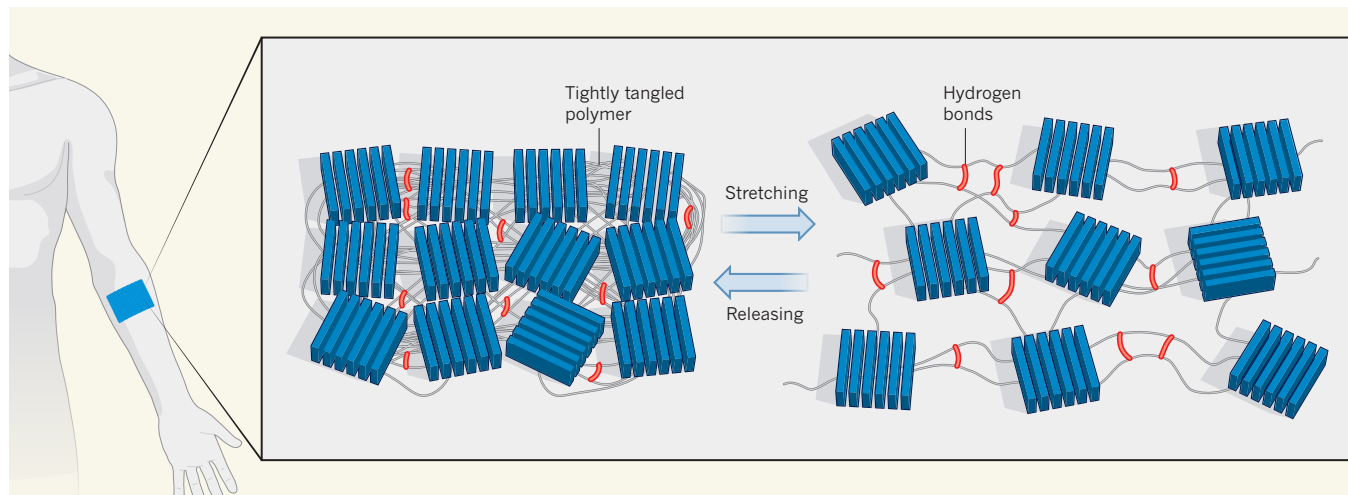


Figure 1 | Molecular design for an intrinsically stretchable and healable organic semiconductor. Oh *et al.*¹ have prepared a polymer that contains semiconducting crystalline domains (blue) and flexible, amorphous polymer chains crosslinked by hydrogen bonds. When the material is subjected to strain, the chains reversibly elongate and the hydrogen-bond network breaks

and re-forms. This allows the material to stretch and unstretch without greatly compromising the semiconducting behaviour. Because broken hydrogen bonds easily re-form, cracks in the bulk material can be healed when heated and treated with solvent vapour. The material was used to make a wearable transistor that tolerated stretching when mounted on human limbs.

In his 1940 book *The Nature of the Chemical Bond and the Structure of Molecules and Crystals*, the chemist Linus Pauling beautifully described the non-covalent origin of hydrogen bonds²: “It has been recognized in recent years that under certain conditions an atom of hydrogen is attracted by rather strong forces to two atoms, instead of only one, so that it may be considered to be acting as a bond between them.” Being weaker than covalent bonds, hydrogen bonds easily rearrange themselves within networks of macromolecules, a fact that has been exploited as a self-healing mechanism in polymers³.

By contrast, hydrogen bonds were thought to be detrimental in organic semiconductors. Research into organic electronics in the 1990s focused mostly on π -conjugated molecules, in which electrons are highly delocalized — an important characteristic for efficient charge transport in a semiconductor. Intermolecular hydrogen bonds were expected to interrupt electron delocalization. But in the past few years, it has emerged that the possible structural basis for building organic semiconductors is much larger than was previously anticipated⁴, and, particularly surprisingly, that charges can move effectively in materials that contain intermolecular hydrogen bonds⁵.

Oh *et al.* have taken that surprising finding to a new level, by incorporating hydrogen bonds into polymers not only to make the materials tolerant to stretching, but also to allow impressive charge transport. They achieved this by tailoring the molecular structure of semiconducting polymers to incorporate crystalline parts, which are responsible for charge transport, and amorphous regions crosslinked through hydrogen bonding, which take up mechanical strain without greatly impairing the material's electronic properties (Fig. 1). The authors

report that films of these polymers crack in response to severe elongation, but that the cracks can be healed by treating the materials with solvent vapour and heat. The healed polymers almost completely recover their initial electronic behaviour.

So far so good, but the next step is to use the materials to make a stretchable electronic device. Three general strategies have been developed to achieve this goal. The first involves placing rigid semiconductors on small flexible ‘islands’ embedded in or placed on elastomers (rubbery materials) that are connected by stretchable wires⁶. This approach also allows high-performance, off-the-shelf

The use of materials that can heal themselves adds a new feature to intrinsically stretchable electronic devices.

microelectronic devices to be integrated onto stretchable materials.

The second strategy is to laminate thin foils containing flexible inorganic⁷ or organic⁸ electronic components onto pre-stretched elastomers. When the elastomer is relaxed, the laminated system wrinkles, thus allowing the device to be subsequently re-stretched in the direction that the elastomer was first extended.

The third approach takes advantage of stretchability at the molecular level, either by using percolating networks of nanostructured conductors or semiconductors — typically tubes or wires — in elastomers⁹, or by incorporating softer materials into a stiff semiconducting polymer without disrupting charge-transport pathways¹⁰. Oh and colleagues' strategy falls into this latter subcategory, but the authors' use of materials that can heal themselves adds a previously

unavailable feature to intrinsically stretchable electronic devices.

Oh *et al.* prepared stretchable transistors from their polymers as proof-of-concept devices, and observed that the devices maintained their electrical performance even after 500 stretching cycles at strains typical of most practical applications. Remarkably, the authors found that the transistors also maintained high charge-carrier mobility when mounted on human limbs and subjected to various common movements, such as twisting of the hand, folding of the arm and stretching of the elbow — as needed for wearable electronic devices.

It should be noted, however, that the authors' transistors are not as durable or stretchable as state-of-the-art devices based on islands and wrinkles, leaving plenty of room for future development. Moreover, the voltage required to operate the new transistors is large (of the order of tens of volts), which lowers the energy efficiency of the devices and is thus a concern for autonomously operating wearables. Smaller voltages will be essential for electronic devices that contact the human body. The voltage problem might be alleviated by greatly reducing the transistors' dimensions — most notably, the thickness of the electrically insulating layer between the control electrode and the transistor's semiconductor — although this would be challenging to achieve. It should also be noted that heating for prolonged periods with solvent vapour may be an impractical means of healing devices worn on human skin.

Perhaps the ultimate goal for scientists developing flexible electronics is to produce something that behaves like human skin: the stretchable organ that covers and protects our bodies, enables us to feel touch, pain and temperature, and triggers a healing process when

wounded. A perfect technological imitation may not be possible, but Oh and colleagues' work is a milestone in the search for electronic skins that behave much like their archetype. In the shorter term, healable soft electronic devices hold promise for truly bionic and smart electrical appliances, and might revolutionize future generations of wearables. ■

Siegfried Bauer and Martin Kaltenbrunner are in the *Soft Matter Physics Department, Johannes Kepler University, Linz 4040, Austria*.
e-mails: sbauer@jku.at;
martin.kaltenbrunner@jku.at

MICROBIOLOGY

Deep-sea secrets of butane metabolism

Anaerobic microbes have been found to break down the hydrocarbon butane by a pathway with some similarities to anaerobic methane breakdown. Harnessing the butane pathway might enable biofuel generation. [SEE ARTICLE P.396](#)

STEPHEN W. RAGSDALE

Microbes living in anaerobic conditions, such as in deep-sea sediments, need a way of generating energy that does not require oxygen gas. Some anaerobic microbes do this by breaking down methane to form carbon dioxide¹. Others² obtain energy by metabolizing the hydrocarbon butane (C₄H₁₀), but the butane-breakdown pathway has until now been unknown. Determining how butane is metabolized might be useful for making biofuels or for developing efficient catalysts for hydrocarbon activation (the process of breaking a hydrocarbon's chemical bonds). On page 396, Laso-Pérez *et al.*³ report the pathway for microbial anaerobic breakdown of butane to CO₂. The authors identify some butane-breakdown steps ripe for further investigation, and observe that some similar steps have evolved in the microbial processes for degrading methane and butane.

Laso-Pérez and colleagues obtained deep-sea sediment samples from the Gulf of California and isolated fractions that could degrade butane. The butane degradation they observed depended on a microbial consortium formed of *Candidatus Syntrophoarchaeum* archaea and HotSeep-1 bacteria, two types of prokaryote (organisms lacking a nucleus). The authors found that the microbial consortium can also metabolize propane (C₃H₈), although it is not known whether propane and butane are

1. Oh, J. Y. *et al.* *Nature* **539**, 411–415 (2016).
2. Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* 2nd edn (Oxford Univ. Press, 1940).
3. Cordier, P., Tournilhac, F., Soulié-Ziakovic, C. & Leibler, L. *Nature* **451**, 977–980 (2008).
4. Gsänger, M., Bialas, D., Huang, L., Stolte, M. & Würthner, F. *Adv. Mater.* **28**, 3615–3645 (2016).
5. Glowacki, E. D. *et al.* *Adv. Mater.* **25**, 1563–1569 (2013).
6. Lacour, S. P., Jones, J., Wagner, S., Li, T. & Suo, Z. *Proc. IEEE* **93**, 1459–1467 (2005).
7. Rogers, J. A., Someya, T. & Huang, Y. *Science* **327**, 1603–1607 (2010).
8. Kaltenbrunner, M. *et al.* *Nature* **499**, 458–463 (2013).
9. Tong, K., Liang, J. & Pei, Q. *ECS Trans.* **75**, 205–212 (2016).
10. Printz, A. D. & Lipomi, D. J. *Appl. Phys. Rev.* **3**, 021302 (2016).

degraded by the same enzymes or by different enzymes that have specific substrate preferences. However, the consortium does not degrade methane, nor several other short-chain hydrocarbons the authors tested.

Butane consists of carbon and hydrogen atoms linked by strong single bonds, and the molecule lacks a reactive group. The stability of the carbon–hydrogen (C–H) bond poses a formidable and interesting chemical challenge in butane breakdown. For anaerobic microbial breakdown of methane, the thermodynamically unfavourable hydrocarbon breakdown proceeds by coupling it to a thermodynamically favourable reaction: for example, the transfer of electrons to sulfate to form hydrogen sulfide (H₂S, a reduced form of sulfate)^{4,5}. Laso-Pérez and colleagues investigated whether such thermodynamic coupling might occur in their microbial consortium, and found evidence for sulfate reduction that depended on butane breakdown. In anaerobic methane degradation, sulfate reduction occurs by direct interspecies electron transfer^{3,4} from the archaea to their bacterial partner HotSeep-1. Laso-Pérez and colleagues find no evidence for sulfate-reduction genes in the archaea they studied, and propose that an interspecies electron-transfer mechanism also occurs in the butane-degradation pathway.

The authors determined experimentally that the overall butane-breakdown reaction consists of the conversion of butane, sulfate



50 Years Ago

Several oncogenic viruses have been known and studied for some years and it is therefore important to search for viruses as aetiological agents in human cancer ... A man 36 years old and of previous good health was referred to one of us ... with clinical evidence of hepatic malignancy. The presence of primary hepatoma had been confirmed at laparotomy. A liver biopsy specimen was obtained by means of a needle ... Histological examination showed the presence of poorly differentiated hepatocarcinoma ... In addition, particles were found in the cytoplasm of some 90 per cent of the eighty tumour cells examined ... Their size and shape indicate that these particles are viral in nature. It has recently been suggested, however, that particles seen as viruses under the electron microscope in various human leukaemia states may have been mycoplasma, small free living micro-organisms with many virus-like properties. The true nature of the particles we have seen must therefore await identification.

From *Nature* 19 November 1966

100 Years Ago

Judging from the correspondence which has recently appeared in *NATURE*, the dearth of wasps this autumn in many parts of England has been most pronounced ... I was staying, with my wife and son, in a cottage on Christon Hill, on the Mendip Hills ... we were simply besieged by wasps, which were particularly tiresome at the breakfast-table. They were so numerous at times ... that we frequently had to abandon our meals temporarily to punish the offenders, my son continuing his captures often for half an hour at a time ... I cannot recollect ever having seen so many wasps in a house, unless it was during the hottest part of 1911.

From *Nature* 16 November 1916

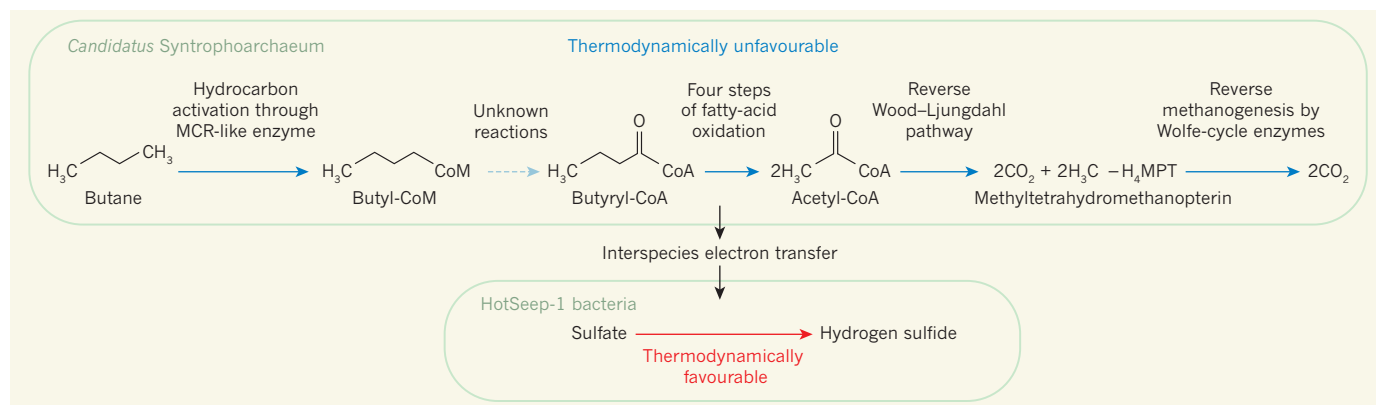


Figure 1 | Proposed pathway for butane breakdown. Laso-Pérez *et al.*³ propose a pathway for the anaerobic breakdown of butane in a microbial consortium formed between *Candidatus Syntrophoarchaeum* archaeal microbes and HotSeep-1 bacteria. Although the overall reaction in *Candidatus Syntrophoarchaeum* is thermodynamically unfavourable, the reaction proceeds by coupling to a thermodynamically favourable reaction in HotSeep-1 bacteria. The initial butane cleavage, a step known as hydrocarbon activation, creates a butane derivative bound to coenzyme M (butyl-CoM). The authors propose that butane activation is catalysed by an enzyme similar to methyl coenzyme M reductase (MCR). Through unknown steps, the intermediate butyl-CoM is converted to butyryl-CoA, which then undergoes

four steps of fatty-acid oxidation to form acetyl-CoA. Wood-Ljungdahl pathway⁷ enzymes convert acetyl-CoA to carbon dioxide and a methyl group bound to the coenzyme tetrahydromethanopterin (H_4MPT). This methyltetrahydromethanopterin derivative is then converted by Wolfe-cycle⁸ enzymes to CO_2 in a process related to the methane degradation pathway known as reverse methanogenesis¹. In the microbial butane-degradation pathway, the Wood-Ljungdahl pathway enzymes and Wolfe-cycle enzymes act to degrade, rather than synthesize, their normal products. These enzyme reversals occur when these reactions are coupled to the thermodynamically favourable reduction of sulfate to hydrogen sulfide, which occurs through interspecies electron transfer to HotSeep-1 bacteria.

and hydrogen ions to final products of CO_2 , H_2S and water. On the basis of clever identification of genes that might have a role in the butane-degradation pathway and determination of possible intermediate compounds and cofactors, Laso-Pérez *et al.* propose a complete degradation pathway from butane to CO_2 in *Candidatus Syntrophoarchaeum*. The authors identify some key stages in the degradation of pathway intermediates (Fig. 1).

The first step in butane degradation involves the breakage of one of butane's C-H single bonds. In methane degradation, C-H bond activation is catalysed by the enzyme methyl coenzyme M reductase (MCR), resulting in the formation of a methyl group bound to coenzyme M (CoM)⁶. Laso-Pérez *et al.* tested whether CoM might also be associated with butane breakdown. Using mass spectrometry, they identified a butane derivative bound to CoM (butyl-CoM) in their microbial sample. The authors also observed that a CoM-mimic molecule, bromoethanesulfonate, inhibits butane degradation. These results strongly implicate the catalytic action of an enzyme similar to MCR in the initial C-H bond cleavage of butane.

The next stage of the breakdown process, the multi-step conversion of butyl-CoM to a derivative bound to coenzyme A (CoA), known as butyryl-CoA, constitutes the most notable mechanistic uncertainty in the pathway described by Laso-Pérez and colleagues. During this sequence, a butane-derived group is transferred from CoM to the ubiquitous CoA. One option is that a butyltransferase enzyme known as Mta catalyses this reaction by transferring the butyl group to CoA, which generates butyl-CoA. However, it seems more

probable that butyl-CoM undergoes oxidation to generate butyryl-CoM, and that the CoM is subsequently displaced by CoA to generate butyryl-CoA.

The authors propose that fatty-acid oxidation steps convert butyryl-CoA to the common metabolic intermediate acetyl-CoA. Then Wood-Ljungdahl pathway⁷ enzymes oxidize acetyl-CoA, resulting in the formation of CO_2 and a methyl group bound to the coenzyme tetrahydromethanopterin (H_4MPT). The final steps of the pathway involve the action of Wolfe-cycle enzymes⁸, which oxidize the methyltetrahydromethanopterin, resulting in the formation of CO_2 .

An enzyme is usually thought of as catalysing a directional reaction that transforms molecule A into molecule B. However, an enzyme actually catalyses a reaction in a state of equilibrium, and if the equilibrium changes, the enzyme can run in reverse and transform molecule B into molecule A. Two of the key metabolic modules in the proposed butane-degradation pathway involve reactions that are best known for running in the opposite direction. In the butane-degradation process, Wood-Ljungdahl enzymes degrade acetyl-CoA, although they usually synthesize acetyl-CoA. Similarly, during butane degradation, Wolfe-cycle enzymes uncharacteristically degrade rather than synthesize a methyl group. This reverse action of Wolfe-cycle enzymes is also observed during methane degradation by anaerobic microbes in the process known as reverse methanogenesis¹.

A particularly interesting enzyme-reversal question arising from this research is whether any 'butanogenic' microbes might naturally exist, or could be engineered, that could run

the entire butane-degradation pathway in reverse, thereby generating butane biofuel. A butane-synthesizing reaction would be thermodynamically favourable, and would not require coupling to sulfate reduction.

Laso-Pérez and colleagues' analysis of the anaerobic consortium that degrades butane provides a glimpse into a pathway for the oxygen-independent breakdown of short-chain hydrocarbons. The authors provide evidence that butane oxidation is linked to sulfate reduction. It will be interesting to investigate whether butane oxidation is coupled to other types of reduction process, such as the reduction of nitrate or metal that occurs in reverse methanogenesis⁹. Further biochemical studies of butane metabolism might lead to major advances in the design of efficient catalysts for hydrocarbon activation, or enable the use of CO_2 to generate butane or other intermediate compounds in this pathway. ■

Stephen W. Ragsdale is in the Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.

e-mail: sragsdal@umich.edu

- Hallam, S. J. *et al. Science* **305**, 1457–1462 (2004).
- Kniemeyer, O. *et al. Nature* **449**, 898–901 (2007).
- Laso-Pérez, R. *et al. Nature* **539**, 396–401 (2016).
- Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. *Nature* **526**, 587–590 (2015).
- McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. *Nature* **526**, 531–535 (2015).
- Scheller, S., Goenrich, M., Boecker, R., Thauer, R. K. & Jaun, B. *Nature* **465**, 606–608 (2010).
- Ragsdale, S. W. & Pierce, E. *Biochim. Biophys. Acta* **1784**, 1873–1898 (2008).
- Thauer, R. K. *Proc. Natl Acad. Sci. USA* **109**, 15084–15085 (2012).
- Joye, S. B. *Nature* **491**, 538–539 (2012).

The effects of Δ^9 -tetrahydrocannabinol on the dopamine system

Michael A. P. Bloomfield^{1,2,3,4,5}, Abhishekh H. Ashok^{1,2,4}, Nora D. Volkow⁶ & Oliver D. Howes^{1,2,4}

The effects of Δ^9 -tetrahydrocannabinol (THC), the main psychoactive ingredient in cannabis, are a pressing concern for global mental health. Patterns of cannabis use are changing drastically owing to legalization, the availability of synthetic analogues (commonly termed spice), cannabavaping and an emphasis on the purported therapeutic effects of cannabis. Many of the reinforcing effects of THC are mediated by the dopamine system. Owing to the complexity of the cannabinoid–dopamine interactions that take place, there is conflicting evidence from human and animal studies concerning the effects of THC on the dopamine system. Acute THC administration causes increased dopamine release and neuron activity, whereas long-term use is associated with blunting of the dopamine system. Future research must examine the long-term and developmental dopaminergic effects of THC.

Cannabis is a widely used recreational drug. Over half of young Americans have used the drug¹, whereas in Europe, cannabis has now overtaken heroin as the most widely reported illegal drug used amongst people entering specialist addiction services². This provides the background to political debates worldwide concerning changes to the legal status of the drug. Although causality has not been conclusively demonstrated, heavy cannabis use is associated with increased risk of mental disorders³ including psychosis⁴, addiction⁵, depression⁶, suicidality⁷, cognitive impairment⁸ and amotivation⁹.

THC, the main psychoactive component of cannabis¹⁰, elicits its acute psychoactive effects through the endocannabinoid type 1 receptor (CB₁R)¹¹. THC has been linked to the rewarding aspects of cannabis and the induction of symptoms of mental illnesses and cognitive impairment. Recently, the THC content of cannabis has been increasing¹² and synthetic THC analogues (potent cannabinoid receptor agonists; termed ‘spice’) are now widely used¹³. The likelihood of a future increase in consumption of cannabinoids through electronic cigarettes (cannabavaping) and edible products¹⁴ changes the landscape further¹⁵. Given the widespread use of cannabinoids, and the links between THC exposure and adverse outcomes, it is imperative that the neurobiological effects of THC are understood. Recently, we and others have found that heavy cannabis use is associated with reductions in dopaminergic function. Since both the rewarding and psychotogenic effects of THC and its analogues are thought to be mediated by the dopaminergic system, demonstrating dopaminergic alterations in human users *in vivo* is of clinical relevance for the prevention and treatment of cannabis-use-associated disorders and psychoses. Therefore, we present a review of the animal and human literature on the complex effects of acute and longer-term THC administration on dopamine synthesis and release and on its receptors, providing an account of the current understanding of the field. We critically analyse the factors that contribute to the effects of THC, and to variations between studies, before providing a framework for future research including a pharmacological dissection of these effects, especially in the developing brain.

THC receptor binding in the brain

THC is a CB₁R and endocannabinoid type 2 receptor (CB₂R) partial agonist¹¹. The psychoactive effects of THC are blocked by the CB₁R antagonist

rimonabant^{16,17}, indicating that these effects are mediated through the activation of G-protein-coupled CB₁R receptors, which reduce cAMP levels by inhibiting adenylate cyclase¹⁸. THC disrupts finely tuned endocannabinoid retrograde signalling systems owing to the temporal and neuronal specificity of endocannabinoids over THC. Under conditions of low CB₁R density, THC antagonizes endogenous agonists that possess greater receptor affinity than THC¹⁹. THC also allosterically modulates opioid receptors²⁰, which may provide additional indirect routes for altering dopamine transmission²¹. Furthermore, THC has psychoactive metabolites with CB₁R affinity, further complicating the analysis of receptor-binding studies²².

CB₁ receptors and dopamine

Early animal studies described the interactions of amphetamine, which increases dopamine release, and THC²³. These studies reported that the behavioural effects of amphetamine were potentiated or antagonized, depending on the dose of THC, leading researchers to propose that dopamine was “a prime candidate for...the mode of action of Δ^9 -tetrahydrocannabinol”²⁴. Indeed, THC produces complex effects on the dopamine system, contributing to the drug’s recreational and harmful effects. However, there are inconsistencies between the preclinical and clinical findings, which have proven challenging for researchers in the field. It is, therefore, timely to review the evidence and provide a framework for understanding the inconsistencies between the preclinical and clinical findings.

Dopaminergic neurons are modulated by the endocannabinoid system²⁵. CB₁Rs and the endocannabinoid ligands anandamide and 2-arachidonoylglycerol (2-AG) are abundant in dopaminergic pathways, including in the striatum²⁶, where they act as a retrograde feedback system on presynaptic glutamatergic and γ -aminobutyric acid (GABA) nerve terminals (Fig. 1), modulating dopamine transmission. Anandamide²⁷ and 2-AG²⁸ stimulate dopamine release in the nucleus accumbens (NAc) shell. Again, this effect is blocked by the CB₁ antagonist rimonabant, indicating that the dopaminergic effects of endocannabinoids involve CB₁ receptors. The rewarding properties of THC, which are mediated through increased dopamine release and dopaminergic neuron firing, are underpinned by biased signal transduction mechanisms from

¹Psychiatric Imaging Group, Robert Steiner MR Unit, MRC Clinical Sciences Centre, Institute of Clinical Sciences, Imperial College London, Hammersmith Hospital, London W12 0NN, UK.

²Psychiatric Imaging Group, Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK. ³Division of Psychiatry, University College London, 6th Floor, Maple House, 149 Tottenham Court Road, London WC1T 7NF, UK. ⁴Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, Kings College London, De Crespigny Park, London SE5 8AF, UK. ⁵Clinical Psychopharmacology Unit, Research Department of Clinical, Educational and Health Psychology, University College London, 1–19 Torrington Place, London WC1E 6BT, UK. ⁶National Institute on Drug Abuse, National Institutes of Health, 6001 Executive Boulevard, Bethesda, Maryland 20892-9561, USA.

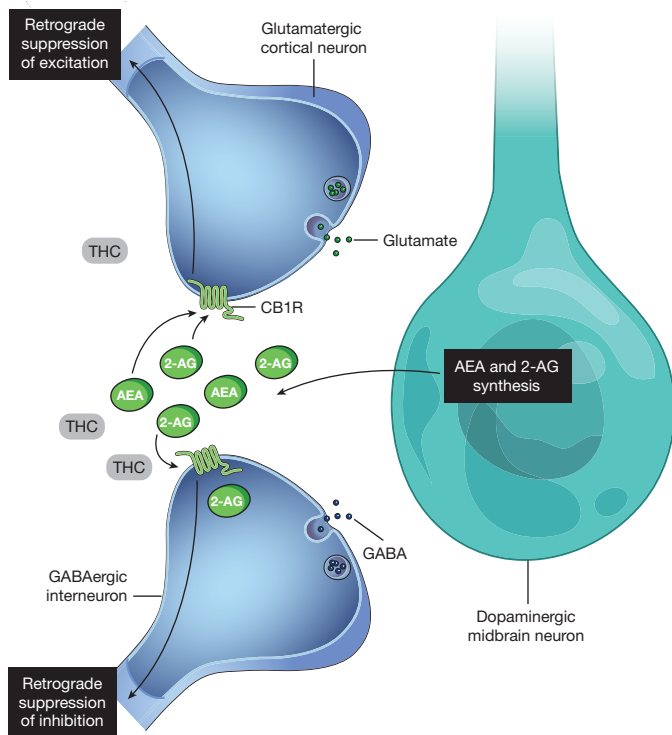


Figure 1 | THC binds to CB1 receptors on glutamatergic and GABAergic neurons, disrupting normal endocannabinoid retrograde signalling from dopaminergic neurons¹³⁵. Endocannabinoids influence synaptic signalling in the midbrain. 2-AG is synthesized by diacylglycerol lipase in dopaminergic midbrain neurons and, once released, acts retroactively on CB₁R on nearby glutamatergic and GABAergic terminals. CB₁R mediate robust inhibition of GABA inputs onto midbrain dopamine cells¹³⁶, termed retrograde suppression of inhibition. CB₁R are also localized on glutamatergic terminals that synapse onto midbrain dopamine neurons¹³⁷, where endocannabinoids mediate retrograde suppression of excitation. Thus, endocannabinoids fine-tune the activity of the mesolimbic dopamine projections through the modulation of both excitatory and inhibitory signalling and THC disrupts this system.

the CB₁R¹⁶. There is evidence that acute and chronic THC exposure have differing effects on the dopaminergic system. We will, therefore, deal with these separately and describe their effects on different neurobiological components of the dopaminergic system, including neuron firing, synthesis, release, reuptake and receptors.

Acute THC and presynaptic dopamine in animals

It has long been clear that THC exerts complex effects on the dopamine system. Early *in vitro* studies in rodent cells, using radiolabelled dopamine in synaptosomes, found that treatment with THC caused increased dopamine synthesis²⁹ and release²⁴ (Fig. 2 and Box 1). However, the effects on dopamine uptake yielded conflicting results, with evidence of both increases³⁰ and dose-dependent decreases²⁴. Subsequently, biphasic and triphasic effects of THC were discovered, whereby low doses of THC produced increases in the conversion of tyrosine to dopamine, but high doses of THC resulted in decreased dopamine synthesis³¹. Similarly complicated temporal relationships between THC administration and changes in dopamine levels were observed³², with repeated dosing resulting in behavioural and neurochemical tolerance—highly pertinent to the mechanisms of dependence to the drug. The complex dose-specific effects of THC in rodents were thought to be due to dose-related decreases in precursor uptake³⁰ and dopamine–opioid interactions via μ -opioid receptors²⁹.

Subsequent work investigated THC-induced increases in dopamine synthesis *in vivo*. THC increased ³H-dopamine synthesis^{33,34}, tyrosine hydroxylase protein levels³⁵ and mRNA expression³⁶, the rate-limiting

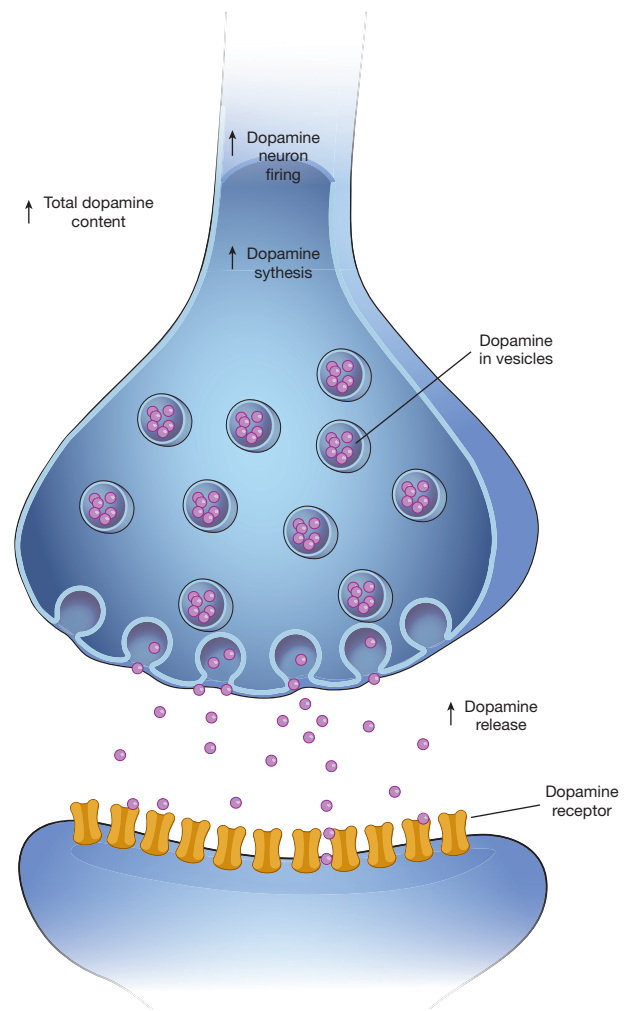


Figure 2 | Summary of the acute effects of THC on dopaminergic function. In animal models, acute THC challenge is associated with increased dopaminergic cell firing, and increased dopamine synthesis and release.

step in the dopamine synthesis pathway. Similarly, increases in dopamine metabolism, measured through the dihydroxyphenylacetic acid/dopamine (DOPAC/DA) ratio, were reported in most³⁷, but not all³⁵, rodent studies. However, the majority of early studies using spectrophotofluorimetry were inconsistent due to technical limitations in detecting the rapid changes in extracellular dopamine concentration, deviations that are now detectable using contemporary microdialysis techniques³⁸.

In vivo microdialysis shows that acute THC administration increases dopamine efflux in the prefrontal cortex (PFC)³⁹, striatum⁴⁰ and NAc⁴¹. Only one study did not find that THC induced increases in dopamine efflux⁴², an anomaly that may have been related to the route of administration as that study used a THC gavage—the other studies used intravenous injection, which produces a rapid increase in THC which reaches the brain promptly, whereas gavage favours sequestration in lipid compartments owing to the very high lipid solubility of THC⁴³. Electrophysiological studies in rats have categorically demonstrated that THC dose-dependently increases firing rates in ascending midbrain dopaminergic projections via CB₁R agonism^{45,46}. Together, these findings suggest that THC increases the firing rates of dopamine neurons, leading to increased dopamine synthesis and release in terminal fields.

Acute THC and postsynaptic dopamine in animals

Acute THC administration did not alter dopamine receptor protein levels in rhesus monkeys⁴⁷. In the rat limbic forebrain, one study reported

BOX 1

Glossary of methods used to assess dopaminergic activity after THC exposure

- High-performance liquid chromatography (HPLC): a technique to separate, identify and quantify a particular chemical from a mixture
- *In vivo* electrophysiology: a method of studying single-neuron responses, using microelectrodes to record from cells, including in the living animal brain
- Microdialysis: a technique used for the continuous measurement of neurochemicals and metabolites through the insertion of a probe into a specific area of the brain
- Positron emission tomography (PET): an imaging technique that uses a positron-emitting radiotracer that binds to specific proteins including receptors, enzymes and transporters. The gamma rays produced by the radionuclide allow quantification of proteins in living tissue non-invasively with very high chemical specificity
- Single-photon emission computed tomography (SPECT): an imaging technique similar to positron emission tomography, but with lower spatial resolution and sensitivity
- Synaptosome: a homogenized mixture of isolated synaptic terminals

increased dopamine type 1 receptor (D₁R) availability⁴⁸ whereas other studies reported decreases⁴⁹. In the striatum, the density of dopamine type 2 receptors (D₂Rs) either decreased⁴⁹ or did not significantly change, with decreases in D₁R also reported⁴⁸. That there is no change in receptor protein levels, together with the tendency towards reduced receptor availability, almost certainly reflects THC-induced changes in synaptic dopamine levels.

Acute THC and dopamine in humans

Studies of brain activity in humans, using functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), provide indirect measures of dopaminergic function through changes in cerebral blood flow and glucose metabolism. These serve as surrogate markers of brain activity in areas with dopaminergic projections. In humans, acute THC administration is associated with increased activity in frontal and sub-cortical regions⁵⁰. However, as the CB₁R has the highest concentrations in these regions⁵¹ these findings may be due to direct effects on the endocannabinoid system rather than representing dopamine-mediated processes. When studies investigating the effects of acute THC administration on resting brain activity have focused on regions with dense dopaminergic innervation, such as the striatum, there have been inconsistent findings, with reports of both increased and reduced activity⁵⁰. However, certain cognitive tasks are modulated by dopaminergic signalling and may provide a more robust proxy for THC-induced changes in dopaminergic transmission. For example, motor response inhibition is associated with cortical dopamine release; an fMRI study of healthy humans with previous exposure to cannabis found that THC attenuates activation in the right inferior frontal cortex and the anterior cingulate cortex (ACC) during suppression of motor inhibition⁵². Further indirect evidence of blunted dopaminergic processing comes from a study of healthy humans with previous exposure to cannabis using a verbal working memory task in which it was found that THC attenuated striatal activation⁵³. Additionally, a study of reward function in occasional cannabis users found that THC induced widespread attenuation of the brain response to feedback in reward trials⁵⁴ but not under reward-anticipation conditions⁵⁴.

It is also possible to directly measure activity of the dopamine system using molecular imaging. These studies have examined the effect of acute THC administration on dopamine release in humans with previous exposure to cannabis *in vivo*. Using PET, a combined analysis⁵⁵ of two previous studies has shown that THC does indeed cause dopamine release in the ventral striatum in the human brain⁵⁵. Likewise, acute THC challenge elicited dopamine release in fronto-temporal cortical brain regions⁵⁶, although this finding requires replication with radiotracers that show higher affinity for cortical dopamine receptors. However, in a separate study using single-photon emission-computed tomography (SPECT), no significant THC-induced dopamine release was observed⁵⁷, which may be due to the relatively small sample size ($n = 9$) and the lower sensitivity of SPECT than PET to dopamine measures. This is likely to be because THC-induced dopamine release in the striatum appears to be of

a lower magnitude than that caused by other drugs such as amphetamine and methylphenidate⁵⁸, combined with difficulties in imaging dopamine changes that are comparatively small⁵⁹.

Repeated THC treatment and presynaptic dopamine

Studies of repeated THC dosing have yielded complex, region-specific effects, such as increased total striatal dopamine levels in rats⁶⁰ but reduced hippocampal dopamine levels in mice⁶¹. Reduced dopamine metabolism was found in the medial PFC in two rat studies⁶⁰, an effect that was not observed in the striatum or the NAc⁶². Administration of THC for 21 days downregulated tyrosine hydroxylase mRNA expression in the substantia nigra and ventral tegmental area (VTA) midbrain nuclei, but not in the cortex or striatum⁶³. Several studies have reported that multiple THC doses do not significantly change dopamine release in the NAc in rats⁶⁴. However, this may be due to differing dopaminergic responses within the NAc; in one rat study, repeated THC doses led to increased dopamine release in the NAc core but decreased release in the NAc shell⁶⁵. The picture is further complicated by genetic strain effects, with increased dopamine release in the NAc observed in Lewis, but not Fischer 344, rat strains⁶⁶. Together, these studies indicate that repeated THC dosing produces region-specific effects on dopamine function.

Studies in human cannabis users

Several molecular imaging studies of dopaminergic function have been conducted in human cannabis users. Using PET, the capacity for dopamine synthesis was observed to be reduced in cannabis users⁶⁷. Notably, this reduction was driven by users meeting clinical criteria for abuse or dependence and was related to the degree of cannabis use (Fig. 3). Similarly, in two separate studies, cannabis users exhibited reduced dopamine release to a stimulant challenge, which was related to degree of cannabis use⁶⁸, and cognitive deficits that included poor working memory⁶⁹. Since no alteration in amphetamine-induced dopamine release was seen in recently abstinent cannabis users⁷⁰, this effect is likely to be related to active use of the drug. Although chronic use was not associated with alterations in stress-induced dopamine release⁷¹, there was evidence of a positive relationship between the duration of cannabis use and the level of stress-induced dopamine release in the limbic striatum⁷¹. Likewise, recent data show that cannabis users have an attenuated metabolic response to methylphenidate challenge in the striatum, with a negative relationship between methylphenidate-induced metabolic increases and severity of cannabis use⁷². There is also evidence of reduced dopamine transporter (DAT) density in chronic cannabis users⁷³. Although the interpretation of some of these studies is complicated by the fact that cannabis users also frequently smoke tobacco, a recent experiment has addressed this by studying cannabis users without comorbid substance dependence, showing that cannabis users do indeed exhibit reduced dopamine release⁶⁹. Overall there is converging evidence that presynaptic dopaminergic function is reduced in cannabis users.

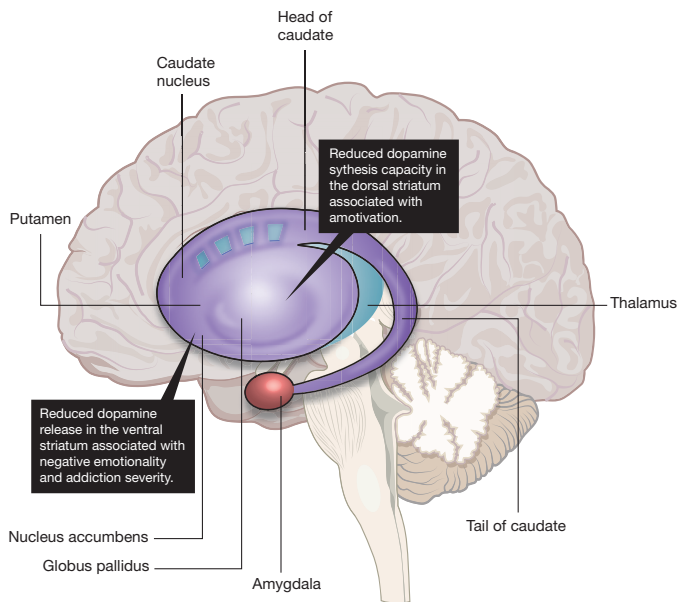


Figure 3 | Cannabis use in humans is associated with reduced dopamine in the striatum. PET studies have shown lower striatal dopamine synthesis and release capacity in cannabis users. Lower dopamine synthesis capacity in the dorsal striatum is directly associated with reduced motivation levels¹⁰², whereas reduced dopamine release in the ventral striatum is directly associated with negative emotion levels and severity of addiction⁶⁸.

Repeated THC administration and postsynaptic dopamine

Studies in rats have reported that multiple THC dosing results in increased D₂R availability in the midbrain, striatum and PFC^{63,74} and that this is associated with dopamine receptor sensitization. There is further evidence of downstream dopaminergic effects of THC; one study in rats⁷⁵ reported upregulated postsynaptic dopamine receptor signalling in the NAc through an increase in adenylyl cyclase activity, which was proposed to underlie THC-induced changes in amphetamine-induced locomotive behaviour. These findings suggest that repeated THC dosing results in altered dopamine receptor signal transduction.

Dose-dependent increases in burst-firing in the VTA in response to multiple THC administrations have been reported in nearly all murine electrophysiological studies⁴⁶, with only one exception of which we are aware⁷⁶. In the substantia nigra pars compacta (SNpc), multiple THC dosing was associated with increased neuronal firing, although this increase was smaller in magnitude than in the VTA^{77,78}. This suggests that VTA and SNpc dopamine neurons develop differing responses to repeated THC exposures. There also appears to be an effect of withdrawal from multiple THC doses, whereby decreased firing is elicited by the abrupt cessation of repeated THC or administration of the CB₁R antagonist SR141716A (ref. 79).

Recent studies in current cannabis users and abstinent former users have found no significant difference in striatal D₂R availability compared to individuals with no history of chronic cannabis use^{68–70,80–82}. However, there was an inverse relationship between the age of first cannabis use and D₂R availability in one study⁷⁰ and an inverse relationship between current cannabis use and receptor availability in another study⁸⁰, suggestive of possible dose effects or susceptibility to drug use.

THC exposure and dopamine cell morphology

In addition to changes in cell function, there is evidence from studies in rodents that THC exposure causes abnormalities in the structure of dopamine neurons. These morphological effects are region-specific and include reductions across a range of measures of neuronal cell size in

the VTA^{83,84} and increased neuronal arborization in the NAc shell and frontal cortex⁸⁵.

Developmental THC exposure and dopamine

Adolescence is an important time for brain development and adolescent cannabinoid exposure has consequences in adulthood⁸⁶. The endocannabinoid system plays an important role in brain development⁸⁷; developmental THC exposure produces complex alterations in the dopamine system that are apparent from an early stage. For example, gestational THC exposure is associated with increased fetal brain tyrosine hydroxylase mRNA expression in rats⁸⁸ and changes in dopamine receptor gene expression. THC exposure in early development is associated with increased cortical D₂R availability⁸⁹. In rats, exposure to THC from gestational day 5 to post-natal day 21 resulted in decreased levels of *DRD2* mRNA (encoding D₂R) in the NAc. Likewise, human maternal cannabis use was associated with decreased *DRD2* mRNA in the fetal NAc⁹⁰. Exposure to THC in early life also blunts the dopaminergic response to dopamine-releasing stimuli, such as stress and amphetamine, later in life. This was seen in the hypothalamus, striatum and limbic forebrain⁹¹, and frontal cortex⁹². In a study of CB₁R agonist self-administration in rats⁹³, THC exposure during adolescence enhanced the reinforcing effects of cannabinoids in adulthood, suggesting that exposure to THC during adolescence increased addiction potential during a critical period of development. Importantly, THC-exposed rodents had a reduced capacity for cannabinoid-induced increases in firing of dopaminergic neurons, consistent with a blunting of the dopamine response in adulthood.

Inconsistencies between animal and human studies

Preclinical evidence shows that acute THC administration increases nerve-firing rates and the results of animal studies using microdialysis indicate that acute THC challenge causes dopamine release, yet the results of experiments from human studies have not been consistent^{94,95}. A number of factors may underlie the inconsistencies in the results of the studies presented above.

The animal evidence indicates that differences in dopamine activity following THC administration are highly region-specific and include differing dopaminergic responses in the shell and core of the NAc⁶⁵. Although human PET imaging provides a reliable measure of dopaminergic function in the striatum⁵⁹, the spatial resolution of most PET cameras is approximately 4 mm, which limits accuracy when measuring activity in small brain volumes (such as in the core and shell of the NAc) owing to partial-volume and spill-over effects⁹⁶. For the foreseeable future, therefore, human *in vivo* imaging techniques lack the spatial resolution to detect these complex effects on different parts of the dopamine pathway, whereas research on living human brain tissue *in vitro* of course raises substantial ethical challenges⁹⁷.

A further possibility underlying the lack of consistent dopamine release in the human studies is that the dose of THC used in human imaging studies has not been sufficient to consistently elicit measurable dopamine release. Human studies all involved the administration of 10 mg or less of THC because older pharmacological studies indicated that the 'standard joint' (or cannabis cigarette; defined by the US National Institute on Drug Abuse) delivered an approximate THC dose of 8–15 mg, equivalent to about 170 µg kg⁻¹ (ref. 98). These doses are considerably less than those used in animal microdialysis studies, which are typically around 1 mg kg⁻¹. Correspondingly, though, the THC content of cannabis has increased considerably since early clinical studies⁹⁹, with current THC doses possibly exceeding 40 mg per joint¹⁰⁰, meaning that the doses that were used in the published imaging studies no longer reflect typical THC exposure in cannabis users. The picture is further complicated by animal data indicating that THC exerts complex dose-dependent effects on the dopamine system³¹; the dose-response profile for THC-induced dopamine release has not been investigated in humans so it is possible that human imaging studies are missing peak dopamine changes. Likewise, heterogeneity in animal data may be a reflection of time of sampling in

relation to THC treatment. For example, repeated daily THC administration led to an initial decrease in dopamine levels, followed by a gradual return to baseline in brains that were studied 1 h after THC dosing, but increases were subsequently observed 2 h after dosing³². This time period is consistent with acute release and diffusion away, followed by an upregulation of synthesis, increasing levels at 2 h. Other factors that may contribute to the complex temporal course of dopaminergic effects include intricate changes in dopamine synthesis and metabolism. These comprise both CB₁R-mediated increases in tyrosine hydroxylase activity³⁶ and acute increases and longer-term decreases in monoamine oxidase activity¹⁰¹, indicating that THC may exert differential effects on the time courses of dopamine synthesis and degradation. Equally, the partial agonist properties of THC on the CB₁R serve only to obfuscate its dopaminergic effects, since THC can both activate and block cannabinoid receptors in a region-specific and species-dependent manner, based on the density and efficiency of populations of receptors and the concentrations of endogenous agonists¹⁹.

There are a number of challenges involved in assessing the long-term effects of exposure to THC. First, it is plausible that discrepant findings may be attributable to the duration of THC treatment. Animal experiments are typically conducted following administration of THC for no longer than three weeks^{63,78,84}, whereas human studies are conducted in participants who have taken cannabis over years (the duration of regular cannabis use is typically in excess of three years¹⁰²). Furthermore, it remains unclear to what degree homeostatic mechanisms are able to compensate for these alterations over time. Additionally, in humans there are a number of psychosocial and genetic risk factors that may predispose an individual to develop a cannabis use disorder. Given the evidence of interactions between early life psychological deprivation and THC-induced effects on dopamine in animals⁷⁴, it is possible that the human data relating to cannabis-dependent participants is confounded by a range of environmental factors, including parental loss and other psychosocial factors¹⁰³. Other environmental and physiological factors have been found to influence the dopaminergic effects of THC. Of particular relevance to humans is sleep deprivation, which decreases dopamine turnover in response to THC¹⁰⁴, and stress, which increases THC-induced dopamine synthesis¹⁰⁵. Cannabis also contains a multitude of other compounds, called phytocannabinoids, which probably exert different complex actions on the dopamine system. These compounds could be behind the varied response observed in human cannabis users when compared to the more consistent animal data that uses only THC. Owing to the co-morbidity of addictions, it is particularly challenging for human imaging researchers to recruit participants with cannabis dependence who do not use other psychoactive compounds, including nicotine and alcohol, in more naturalistic studies.

Finally, whether THC is delivered in a contingent or non-contingent way is likely to be important, given evidence that anticipation leads to dopamine release. In general, animal studies include a variety of contingencies around the dosing of THC. By contrast, in human acute studies THC was given to participants with a prior exposure to the drug but was not delivered in the habitual manner in which the drug is consumed, potentially reducing the degree of dopamine release.

Dopamine and the behavioural effects of THC

THC is associated with a number of behavioural effects that are likely to involve alterations in dopaminergic function. The first evidence of this came from early animal studies, which reported similarities between the behavioural effects of THC, including catalepsy and hypothermia¹⁰⁶, and dopamine antagonists¹⁰⁷. However, THC also antagonizes the locomotor and hyperthermic effects of amphetamine, which causes dopamine release¹⁰⁸. These seemingly paradoxical effects may be due to the high doses used in the early research together with the partial agonist effects of THC and/or non-specific receptor-independent effects.

THC promotes increased food intake in animals¹⁰⁹ and humans¹¹⁰. Since appetite is modulated by the dopamine system¹¹¹, studies have

investigated dopaminergic involvement in THC-induced feeding. The dopamine D₁ receptor antagonist SCH23390 attenuated THC-induced feeding at a dose that did not affect feeding on its own¹¹². Recent research has identified CB₁R-mediated changes in hypothalamic pro-opiomelanocortin (POMC) neurons¹¹³ as the potential underlying cause of this process, through mitochondrial uncoupling protein 2, which is involved in ghrelin-mediated dopaminergic function¹¹⁴.

Heavy cannabis use is associated with impaired educational and occupational outcomes¹¹⁵. Factors that may underlie this association include cognitive impairment, which includes executive dysfunction¹¹⁶, working memory impairments¹¹⁷ and amotivation¹¹⁸ (defined as reduced motivation for goal-directed behaviour¹¹⁹). These functions are susceptible to mesocortical dopaminergic manipulation¹²⁰ (for example, prefrontal D₁ receptor blockade¹²¹). Although a preclinical study reported that D₂ receptor antagonism blocks THC-induced working memory deficits¹²², this was not replicated in humans¹²³. Nonetheless, there is recent evidence that THC-induced working-memory deficits are moderated by catechol-*O*-methyltransferase, a key enzyme in the dopamine metabolic pathway¹²⁴, which also modulates the effects of THC use on dopaminergic cell size in adolescents⁸⁴. Chronic heavy cannabis use produces apathetic behaviour in rhesus monkeys¹²⁵, while a study in humans¹⁰² found that the reduced dopamine synthesis capacity observed in heavy cannabis users was related to amotivation. Overlapping features of the amotivational syndrome associated with cannabis use disorders, such as reduced reward sensitivity and negative emotionality¹²⁶, were also found to be inversely related to methylphenidate-induced dopamine ventral striatal dopamine release⁶⁸.

The dopamine system is involved in risk of psychosis¹²⁷. An early case report described increased striatal dopamine release following cannabis intoxication, associated with the exacerbation of psychotic symptoms in a patient with schizophrenia¹²⁸. Furthermore, cannabis users with a diagnosis of schizophrenia, and those at clinical high risk for schizophrenia, displayed blunted striatal stress-induced dopamine release¹²⁹. Although dopamine release was blunted in cannabis users with schizophrenia, it was nevertheless directly related to the induction of psychotic symptoms¹³⁰. Supersensitivity of postsynaptic D₂Rs⁶³ could explain this apparent paradox; alternatively it could be due to impaired endocannabinoid regulation of dopamine signal transduction. Patients with schizophrenia using high levels of cannabis show reduced anandamide levels, supporting the latter explanation¹³¹. Accordingly, cannabidiol, a phytocannabinoid compound that elevates anandamide levels, has been shown to reduce psychotic symptoms¹³². Alternative postsynaptic mechanisms include CB₁R–D₂R heterodimerization¹³³ and downstream intracellular mechanisms such as the neuregulin 1–erbB2 receptor tyrosine kinase 4–phosphoinositide 3 kinase–protein kinase B (NRG1–ERBB4–PI3K–AKT1) pathway¹³⁴.

Outlook

There is now substantial evidence that THC exerts effects on the dopamine system in animals. The key challenges for the field must be to understand the complexity of these effects, identify how they translate to humans and appreciate how they relate to the potential negative effects of the drug in humans (Box 2). Animal studies demonstrate that the acute administration of THC causes region-specific increases in dopamine release and nerve activity—we must understand the functional significance of these increases. The available preclinical evidence suggests that chronic THC administration causes long-term changes to the dopamine system, but these models are limited as they do not reflect typical patterns of human use. Future studies should, therefore, be of longer duration to reflect human use and should also include co-administration with other drugs, such as nicotine and alcohol, which are commonly used with cannabis. Related to this is a need to understand how the dopaminergic effects of THC are modulated by the other phytocannabinoids. Human PET studies have demonstrated that dopamine synthesis and dopamine release are blunted in cannabis

BOX 2

Strategy for future work

Preclinical and clinical research

- Identify the relationships between cannabis-induced alterations in the dopamine system and behavioural phenomena in humans and animal models
- Use translational techniques that can be applied in human and animal studies alike and employ study designs that better reflect patterns of human use, including modelling contingency in acute THC challenge studies
- Achieve consensus on dose equivalence across species
- Determine whether THC-induced dopaminergic changes during key developmental phases persist into later life and whether this is linked to behavioural changes
- Investigate how gene variants that modulate the endocannabinoid and dopamine systems influence the sensitivity to the rewarding effects of THC and vulnerability to addiction, amotivation and psychosis following chronic exposure
- Investigate the effects of sex in THC response. This will be useful to understand the differences in susceptibility to cannabis effects that have been reported between males and females⁷².

Preclinical research

- Identify the effects on the dopamine system of long-term THC exposure alongside co-administration with nicotine to reflect typical patterns of human use
- Determine the mechanisms underlying the complex dose–response effects of THC on dopaminergic function
- Elucidate the mechanisms of regional differences in dopaminergic effects and the functional significance of these on behaviour
- Determine if the long-term effects of THC are reversible with abstinence

Clinical research

- Determine if the blunted dopamine release and synthesis seen in chronic users is a pre-existing vulnerability factor or a direct result of repeated THC exposure
- Determine the dopaminergic changes over the course of repeated THC exposures and dose–response effects
- Achieve consensus on how to report previous exposure to cannabis use in the human literature
- Determine whether there are regional differences in dopamine release in response to THC administration in humans
- Determine the dose and context dependency of THC effects on the dopamine system

users relative to non-users, yet we still need to understand the precise mechanisms through which this occurs.

It is necessary to identify whether the effects of THC on the dopamine system are reversible and, if they are not, to recognize when they become so. Likewise, further molecular imaging studies in humans are needed to determine the dose–response and timing of the acute effects of THC on dopamine release, and to determine whether there are regional differences in dopamine release—particularly between cortical and sub-cortical regions.

Ultimately, one key challenge in understanding the effects of THC on the dopamine system is translating results from animal studies, many of which cannot be carried out clinically, into humans. We must also reach consensus on THC dose equivalence across species. It is therefore difficult to know what the implications of some preclinical animal work are for human research and how to back-translate human findings into preclinical models. There must be movement away from untranslatable, standalone research, towards the use of translational techniques, such as PET and MRI approaches that can be conducted in both humans and preclinical models that can capture human use patterns in combination with traditional preclinical techniques. We must also understand how psychoactive metabolites of THC and endocannabinoids modulate THC-induced changes in dopaminergic function.

Given the changing patterns in cannabis use across the world, particularly in young people, and the increase in the THC content of cannabis, there is clearly a pressing need to understand how THC alters dopaminergic function. This is especially important given that emerging evidence that links dopaminergic alterations to a number of the adverse cognitive and behavioural consequences of THC, and that there is a dearth of current effective biological interventions for many of the psychiatric sequelae of cannabis use. Dopaminergic dysfunction may thus represent an area for future treatment targets.

There is evidence that gestational exposure to THC is associated with dysregulated dopamine synthesis in later life. This has major potential public health implications, given the prevalence of cannabis use in women of child-bearing age and that the liberalization of cannabis laws around the world may be associated with increased use of the drug amongst gravid and nursing mothers. However, there remain gaps in our knowledge of the developmental effects of THC. In particular, we do not know how long the dopaminergic effects persist for and what the behavioural consequences are. Perhaps most critically, we do not yet know what the effects of THC exposure are on the adolescent dopamine system.

The available evidence indicates that THC exposure produces complex, diverse and potentially long-term effects on the dopamine system. These include increased nerve firing and dopamine release in response to acute THC, and dopaminergic blunting associated with long-term use. Future research should focus on probing the relationships between cannabis-induced alterations in the dopamine system and behavioural effects in humans and animal models.

Received 3 March; accepted 10 October 2016.

1. Substance Abuse and Mental Health Services Administration. *Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings* Vol. HHS Publication No. (SMA) 14–4863 (Substance Abuse and Mental Health Services Administration, 2014).
2. EMCDDA. *European Drug Report 2015: Trends and Developments* (European Monitoring Centre for Drugs and Drug Addiction, Lisbon, 2015).
3. Volkow, N. D. *et al.* Effects of cannabis use on human behavior, including cognition, motivation, and psychosis: a review. *JAMA Psychiatry* **73**, 292–297 (2016).
4. Di Forti, M. *et al.* Daily use, especially of high-potency cannabis, drives the earlier onset of psychosis in cannabis users. *Schizophr. Bull.* **40**, 1509–1517 (2014).
5. Fergusson, D. M., Boden, J. M. & Horwood, L. J. Cannabis use and other illicit drug use: testing the cannabis gateway hypothesis. *Addiction* **101**, 556–569 (2006).
6. Horwood, L. J. *et al.* Cannabis and depression: an integrative data analysis of four Australasian cohorts. *Drug Alcohol Depend.* **126**, 369–378 (2012).
7. Silins, E. *et al.* Young adult sequelae of adolescent cannabis use: an integrative analysis. *Lancet Psychiatry* **1**, 286–293 (2014).
8. Crane, N. A., Schuster, R. M., Fusar-Poli, P. & Gonzalez, R. Effects of cannabis on neurocognitive functioning: recent advances, neurodevelopmental influences, and sex differences. *Neuropsychol. Rev.* **23**, 117–137 (2013).
9. Cherek, D. R., Lane, S. D. & Dougherty, D. M. Possible amotivational effects following marijuana smoking under laboratory conditions. *Exp. Clin. Psychopharmacol.* **10**, 26–38 (2002).
10. Wachtel, S. R., ElSohly, M. A., Ross, S. A., Ambre, J. & de Wit, H. Comparison of the subjective effects of Δ^9 -tetrahydrocannabinol and marijuana in humans. *Psychopharmacology (Berl.)* **161**, 331–339 (2002).
11. Felder, C. C., Veluz, J. S., Williams, H. L., Briley, E. M. & Matsuda, L. A. Cannabinoid agonists stimulate both receptor- and non-receptor-mediated signal transduction pathways in cells transfected with and expressing cannabinoid receptor clones. *Mol. Pharmacol.* **42**, 838–845 (1992).
12. Mehmedic, Z. *et al.* Potency trends of Δ^9 -THC and other cannabinoids in confiscated cannabis preparations from 1993 to 2008. *J. Forensic Sci.* **55**, 1209–1217 (2010).
13. Spaderna, M., Addy, P. H. & D'Souza, D. C. Spicing things up: synthetic cannabinoids. *Psychopharmacology (Berl.)* **228**, 525–540 (2013).
14. Benjamin, D. M. & Fossler, M. J. Edible cannabis products: it is time for FDA oversight. *J. Clin. Pharmacol.* (2016).

15. Varlet, V. *et al.* Drug vaping applied to cannabis: Is “Cannavaping” a therapeutic alternative to marijuana? *Sci. Rep.* **6**, 25599 (2016).
16. Vallee, M. *et al.* Pregnenolone can protect the brain from cannabis intoxication. *Science* **343**, 94–98 (2014).
17. Huestis, M. A. *et al.* Blockade of effects of smoked marijuana by the CB1-selective cannabinoid receptor antagonist SR141716. *Arch. Gen. Psychiatry* **58**, 322–328 (2001).
- The psychoactive active effects of cannabis are mediated via the CB₁ receptor.**
18. Elphick, M. R. & Egertova, M. The neurobiology and evolution of cannabinoid signalling. *Phil. Trans. R. Soc. Lond. B* **356**, 381–408 (2001).
19. Pertwee, R. G. The diverse CB₁ and CB₂ receptor pharmacology of three plant cannabinoids: Δ^9 -tetrahydrocannabinol, cannabidiol and Δ^9 -tetrahydrocannabinol. *Br. J. Pharmacol.* **153**, 199–215 (2008).
20. Kathmann, M., Flau, K., Redmer, A., Trankle, C. & Schlicker, E. Cannabidiol is an allosteric modulator at μ - and δ -opioid receptors. *Naunyn Schmiedeberg's Arch. Pharmacol.* **372**, 354–361 (2006).
21. Chartoff, E. H. & Connery, H. S. It's MORe exciting than mu: crosstalk between mu opioid receptors and glutamatergic transmission in the mesolimbic dopamine system. *Front. Pharmacol.* **5**, 116 (2014).
22. Hollister, L. E. & Gillespie, H. K. Action of Δ^9 -tetrahydrocannabinol. An approach to the active metabolite hypothesis. *Clin. Pharmacol. Ther.* **18**, 714–719 (1975).
23. Garriott, J. C., King, L. J., Forney, R. B. & Hughes, F. W. Effects of some tetrahydrocannabinols on hexobarbital sleeping time and amphetamine induced hyperactivity in mice. *Life Sci.* **6**, 2119–2128 (1967).
24. Howes, J. & Osgood, P. The effect of Δ^9 -tetrahydrocannabinol on the uptake and release of 14 C-dopamine from crude striatal synaptosomes; preparations. *Neuropharmacology* **13**, 1109–1114 (1974).
25. Fernandez-Ruiz, J., Hernandez, M. & Ramos, J. A. Cannabinoid-dopamine interaction in the pathophysiology and treatment of CNS disorders. *CNS Neurosci. Ther.* **16**, e72–e91 (2010).
26. Herkenham, M., Lynn, A. B., Decosta, B. R. & Richfield, E. K. Neuronal localization of cannabinoid receptors in the basal ganglia of the rat. *Brain Res.* **547**, 267–274 (1991).
27. Solinas, M., Justinova, Z., Goldberg, S. R. & Tanda, G. Anandamide administration alone and after inhibition of fatty acid amide hydrolase (FAAH) increases dopamine levels in the nucleus accumbens shell in rats. *J. Neurochem.* **98**, 408–419 (2006).
28. De Luca, M. A. *et al.* Endocannabinoid 2-arachidonoylglycerol self-administration by Sprague–Dawley rats and stimulation of *in vivo* dopamine transmission in the nucleus accumbens shell. *Front. Psychiatry* **5**, 140 (2014).
29. Bloom, A. S. & Dewey, W. L. A comparison of some pharmacological actions of morphine and Δ^9 -tetrahydrocannabinol in the mouse. *Psychopharmacology (Berl.)* **57**, 243–248 (1978).
30. Hershkowitz, M. & Szechtman, H. Pretreatment with Δ^1 -tetrahydrocannabinol and psychoactive drugs: effects on uptake of biogenic amines and on behavior. *Eur. J. Pharmacol.* **59**, 267–276 (1979).
31. Poddar, M. K. & Dewey, W. L. Effects of cannabinoids on catecholamine uptake and release in hypothalamic and striatal synaptosomes. *J. Pharmacol. Exp. Ther.* **214**, 63–67 (1980).
32. Aulakh, C. S., Bhattacharyya, A. K., Hossain, M. A. & Pradhan, S. N. Behavioral and neurochemical effects of repeated administration of Δ^9 -tetrahydrocannabinol in rats. *Neuropharmacology* **19**, 97–102 (1980).
33. Maitre, L., Staehelin, M. & Bein, H. J. Effect of an extract of cannabis and of some cannabinoids on catecholamine metabolism in rat brain and heart. *Agents Actions* **1**, 136–143 (1970).
34. Bloom, A. S., Johnson, K. M. & Dewey, W. L. The effects of cannabinoids on body temperature and brain catecholamine synthesis. *Res. Commun. Chem. Pathol. Pharmacol.* **20**, 51–57 (1978).
35. Romero, J., Demiguel, R., Garciapalmero, E., Fernandezruiz, J. J. & Ramos, J. A. Time-course of the effects of anandamide, the putative endogenous cannabinoid receptor-ligand, on extrapyramidal function. *Brain Res.* **694**, 223–232 (1995).
36. Bosier, B. *et al.* Differential modulations of striatal tyrosine hydroxylase and dopamine metabolism by cannabinoid agonists as evidence for functional selectivity *in vivo*. *Neuropharmacology* **62**, 2328–2336 (2012).
37. Navarro, M. *et al.* An acute dose of Δ^9 -tetrahydrocannabinol affects behavioral and neurochemical indices of mesolimbic dopaminergic activity. *Behav. Brain Res.* **57**, 37–46 (1993).
38. Heien, M. L. *et al.* Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. *Proc. Natl Acad. Sci. USA* **102**, 10023–10028 (2005).
39. Pistis, M. *et al.* Δ^9 -Tetrahydrocannabinol decreases extracellular GABA and increases extracellular glutamate and dopamine levels in the rat prefrontal cortex: An *in vivo* microdialysis study. *Brain Res.* **948**, 155–158 (2002).
40. Ng Cheong Ton, J. M. *et al.* The effects of Δ^9 -tetrahydrocannabinol on potassium-evoked release of dopamine in the rat caudate nucleus: an *in vivo* electrochemical and *in vivo* microdialysis study. *Brain Res.* **451**, 59–68 (1988).
41. Chen, J. P. *et al.* Δ^9 -tetrahydrocannabinol produces naloxone-blockable enhancement of presynaptic basal dopamine efflux in nucleus accumbens of conscious, freely-moving rats as measured by intracerebral microdialysis. *Psychopharmacology (Berl.)* **102**, 156–162 (1990).
42. Castaneda, E., Moss, D., Oddie, S. & Whishaw, I. THC does not affect striatal dopamine release: Microdialysis in freely moving rats. *Pharmacol. Biochem. Behav.* **40**, 587–591 (1991).
43. Nahas, G. G. in *Medical Aspects of Drug Abuse* (ed. R. W. Richter) 16–36 (Harper & Row, 1975).
44. Chen, J., Paredes, W., Lowinson, J. & Gardner, E. Strain-specific facilitation of dopamine efflux by Δ^9 -tetrahydrocannabinol in the nucleus accumbens of rat: An *in vivo* microdialysis study. *Neurosci. Lett.* **129**, 136–140 (1991).
45. Oleson, E. B. & Cheer, J. F. A brain on cannabinoids: the role of dopamine release in reward seeking. *Cold Spring Harb. Perspect. Med.* **2**, a012229 (2012).
46. French, E. Δ^9 -Tetrahydrocannabinol excites rat VTA dopamine neurons through activation of cannabinoid CB1 but not opioid receptors. *Neurosci. Lett.* **226**, 159–162 (1997).
47. Ali, S. F. *et al.* Chronic marijuana smoke exposure in the rhesus monkey. IV: Neurochemical effects and comparison to acute and chronic exposure to Δ^9 -tetrahydrocannabinol (THC) in rats. *Pharmacol. Biochem. Behav.* **40**, 677–682 (1991).
48. Navarro, M. *et al.* Motor disturbances induced by an acute dose of Δ^9 -tetrahydrocannabinol: Possible involvement of nigrostriatal dopaminergic alterations. *Pharmacol. Biochem. Behav.* **45**, 291–298 (1993).
49. Rodríguez De Fonseca, F. *et al.* Acute effects of Δ^9 -tetrahydrocannabinol on dopaminergic activity in several rat-brain areas. *Pharmacol. Biochem. Behav.* **42**, 269–275 (1992).
50. Volkow, N. D. *et al.* Brain glucose metabolism in chronic marijuana users at baseline and during marijuana intoxication. *Psychiatry Res.* **67**, 29–38 (1996).
51. Pertwee, R. G. & Ross, R. A. Cannabinoid receptors and their ligands. *Prostaglandins Leukot. Essent. Fatty Acids* **66**, 101–121 (2002).
52. Borgwardt, S. J. *et al.* Neural basis of Δ^9 -tetrahydrocannabinol and cannabidiol: effects during response inhibition. *Biol. Psychiatry* **64**, 966–973 (2008).
53. Bhattacharyya, S. *et al.* Opposite effects of Δ^9 -tetrahydrocannabinol and cannabidiol on human brain function and psychopathology. *Neuropsychopharmacology* **35**, 764–774 (2010).
54. van Hell, H. H. *et al.* Involvement of the endocannabinoid system in reward processing in the human brain. *Psychopharmacology (Berl.)* **219**, 981–990 (2012).
55. Bossong, M. G. *et al.* Further human evidence for striatal dopamine release induced by administration of Δ^9 -tetrahydrocannabinol (THC): selectivity to limbic striatum. *Psychopharmacology (Berl.)* (2015).
56. Stokes, P. R. *et al.* Significant decreases in frontal and temporal [11C]-raclopride binding after THC challenge. *Neuroimage* **52**, 1521–1527 (2010).
57. Barkus, E. *et al.* Does intravenous Δ^9 -tetrahydrocannabinol increase dopamine release? A SPET study. *J. Psychopharmacol.* **25**, 1462–1468 (2011).
58. Volkow, N. D. *et al.* Reinforcing effects of psychostimulants in humans are associated with increases in brain dopamine and occupancy of D₂ receptors. *J. Pharmacol. Exp. Ther.* **291**, 409–415 (1999).
59. Egerton, A., Demjaha, A., McGuire, P., Mehta, M. A. & Howes, O. D. The test-retest reliability of 18F-DOPA PET in assessing striatal and extrastriatal presynaptic dopaminergic function. *Neuroimage* **50**, 524–531 (2010).
60. Jentsch, J., Verrico, C., Le, D. & Roth, R. Repeated exposure to Δ^9 -tetrahydrocannabinol reduces prefrontal cortical dopamine metabolism in the rat. *Neurosci. Lett.* **246**, 169–172 (1998).
61. Avraham, Y. *et al.* Very low doses of Δ^9 -THC increase food consumption and alter neurotransmitter levels following weight loss. *Pharmacol. Biochem. Behav.* **77**, 675–684 (2004).
62. Jentsch, J., Andrusiak, E., Tran, A., Bowers, Jr. & Roth, R. Δ^9 -Tetrahydrocannabinol increases prefrontal cortical catecholaminergic utilization and impairs spatial working memory in the rat: Blockade of dopaminergic effects with HA966. *Neuropsychopharmacology* **16**, 426–432 (1997).
63. Ginovart, N. *et al.* Chronic Δ^9 -tetrahydrocannabinol exposure induces a sensitization of dopamine D_{2/3} receptors in the mesoaccumbens and nigrostriatal systems. *Neuropsychopharmacology* **37**, 2355–2367 (2012).
64. Tanda, G., Pontieri, F. E. & Di Chiara, G. Cannabinoid and heroin activation of mesolimbic dopamine transmission by a common μ 1 opioid receptor mechanism. *Science* **276**, 2048–2050 (1997).
- THC increased extracellular dopamine concentrations in the nucleus accumbens shell.**
65. Cadoni, C., Valentini, V. & Di Chiara, G. Behavioral sensitization to Δ^9 -tetrahydrocannabinol and cross-sensitization with morphine: differential changes in accumbal shell and core dopamine transmission. *J. Neurochem.* **106**, 1586–1593 (2008).
66. Cadoni, C., Simola, N., Espa, E., Fenu, S. & Di Chiara, G. Strain dependence of adolescent cannabis influence on heroin reward and mesolimbic dopamine transmission in adult Lewis and Fischer 344 rats. *Addict. Biol.* **20**, 132–142 (2015).
67. Bloomfield, M. A. P. *et al.* Dopaminergic function in cannabis users and its relationship to cannabis-induced psychotic symptoms. *Biol. Psychiatry* **75**, 470–478 (2014).
- Dopamine synthesis capacity is reduced in long-term human cannabis users.**
68. Volkow, N. D. *et al.* Decreased dopamine brain reactivity in marijuana abusers is associated with negative emotionality and addiction severity. *Proc. Natl Acad. Sci. USA* **111**, E3149–E3156 (2014).
- Dopamine release is blunted in chronic human cannabis users.**
69. van de Giessen, E. *et al.* Deficits in striatal dopamine release in cannabis dependence. *Mol. Psychiatry* <http://dx.doi.org/10.1038/mp.2016.21> (2016).

70. Urban, N. B. L. *et al.* Dopamine release in chronic cannabis users: a [11 C]-raclopride positron emission tomography study. *Biol. Psychiatry* **71**, 677–683 (2012).
71. Mizrahi, R. *et al.* Dopamine response to psychosocial stress in chronic cannabis users: a PET study with [11 C]-(-)-PHNO. *Neuropsychopharmacology* **38**, 673–682 (2013).
72. Wiers, C. E. *et al.* Cannabis abusers show hypofrontality and blunted brain responses to a stimulant challenge in females but not in males. *Neuropsychopharmacology* **41**, 2596–2605 (2016).
73. Leroy, C. *et al.* Striatal and extrastriatal dopamine transporter in cannabis and tobacco addiction: a high-resolution PET study. *Addict. Biol.* **17**, 981–990 (2012).
74. Zamberletti, E. *et al.* Gender-dependent behavioral and biochemical effects of adolescent Δ^9 -tetrahydrocannabinol in adult maternally deprived rats. *Neuroscience* **204**, 245–257 (2012).
75. Cortright, J. J., Lorrain, D. S., Beeler, J. A., Tang, W. J. & Vezina, P. Previous exposure to Δ^9 -tetrahydrocannabinol enhances locomotor responding to but not self-administration of amphetamine. *J. Pharmacol. Exp. Ther.* **337**, 724–733 (2011).
76. Gifford, A., Gardner, E. & Ashby Jr., C. R. The effect of intravenous administration of Δ^9 -tetrahydrocannabinol on the activity of A10 dopamine neurons recorded in vivo in anesthetized rats. *Neuropsychobiology* **36**, 96–99 (1997).
77. French, E. D., Dillon, K. & Wu, X. Cannabinoids excite dopamine neurons in the ventral tegmentum and substantia nigra. *Neuroreport* **8**, 649–652 (1997).
78. Wu, X. & French, E. Effects of chronic Δ^9 -tetrahydrocannabinol on rat midbrain dopamine neurons: An electrophysiological assessment. *Neuropharmacology* **39**, 391–398 (2000).
79. Diana, M., Melis, M., Muntoni, A. L. & Gessa, G. L. Mesolimbic dopaminergic decline after cannabinoid withdrawal. *Proc. Natl Acad. Sci. USA* **95**, 10269–10273 (1998).
- Chronic THC exposure is associated with reduced dopamine neuron activity in the meso-accumbens.**
80. Albrecht, D. S. *et al.* Striatal D₂/D₃ receptor availability is inversely correlated with cannabis consumption in chronic marijuana users. *Drug Alcohol Depend.* **128**, 52–57 (2013).
81. Sevy, S. *et al.* Cerebral glucose metabolism and D₂/D₃ receptor availability in young adults with cannabis dependence measured with positron emission tomography. *Psychopharmacology (Berl.)* **197**, 549–556 (2008).
82. Stokes, P. R. A. *et al.* History of cannabis use is not associated with alterations in striatal dopamine D₂/D₃ receptor availability. *J. Psychopharmacol.* **26**, 144–149 (2012).
83. Spiga, S., Lintas, A., Migliore, M. & Diana, M. Altered architecture and functional consequences of the mesolimbic dopamine system in cannabis dependence. *Addict. Biol.* **15**, 266–276 (2010).
84. Behan, A. *et al.* Chronic adolescent exposure to Δ^9 -tetrahydrocannabinol in COMT mutant mice: Impact on indices of dopaminergic, endocannabinoid and GABAergic pathways. *Neuropsychopharmacology* **37**, 1773–1783 (2012).
85. Kolb, B., Gorny, G., Limebeer, C. L. & Parker, L. A. Chronic treatment with Δ^9 -tetrahydrocannabinol alters the structure of neurons in the nucleus accumbens shell and medial prefrontal cortex of rats. *Synapse* **60**, 429–436 (2006).
86. Renard, J., Krebs, M. O., Le Pen, G. & Jay, T. M. Long-term consequences of adolescent cannabinoid exposure in adult psychopathology. *Front. Neurosci.* **8**, 361 (2014).
87. Berghuis, P. *et al.* Hardwiring the brain: endocannabinoids shape neuronal connectivity. *Science* **316**, 1212–1216 (2007).
88. Bonnin, A., de Miguel, R., Hernandez, M. L., Ramos, J. A. & Fernandez-Ruiz, J. J. The prenatal exposure to Δ^9 -tetrahydrocannabinol affects the gene expression and the activity of tyrosine hydroxylase during early brain development. *Life Sci.* **56**, 2177–2184 (1995).
89. Walters, D. E. & Carr, L. A. Perinatal exposure to cannabinoids alters neurochemical development in rat brain. *Pharmacol. Biochem. Behav.* **29**, 213–216 (1988).
90. DiNieri, J. A. *et al.* Maternal cannabis use alters ventral striatal dopamine D₂ gene regulation in the offspring. *Biol. Psychiatry* **70**, 763–769 (2011).
- Prenatal cannabis exposure decreases dopamine receptor D₂ messenger RNA expression in the ventral striatum of offspring in humans.**
91. Garcia-Gil, L. *et al.* Perinatal Δ^9 -tetrahydrocannabinol exposure alters the responsiveness of hypothalamic dopaminergic neurons to dopamine-acting drugs in adult rats. *Neurotoxicol. Teratol.* **19**, 477–487 (1997).
92. Mokler, D. J., Robinson, S. E., Johnson, J. H., Hong, J. S. & Rosecrans, J. A. Neonatal administration of Δ^9 -tetrahydrocannabinol (THC) alters the neurochemical response to stress in the adult Fischer-344 rat. *Neurotoxicol. Teratol.* **9**, 321–327 (1987).
93. Scherma, M. *et al.* Adolescent Δ^9 -tetrahydrocannabinol exposure alters WIN55,212-2 self-administration in adult rats. *Neuropsychopharmacology* **41**, 1416–1426 (2016).
94. Bossong, M. G. *et al.* Δ^9 -tetrahydrocannabinol induces dopamine release in the human striatum. *Neuropsychopharmacology* **34**, 759–766 (2009).
- Combined analysis of two previous PET studies showing that acute THC administration causes dopamine release in humans.**
95. Stokes, P., Mehta, M., Curran, H., Breen, G. & Grasby, P. Can recreational doses of THC produce significant dopamine release in the human striatum? *Neuroimage* **48**, 186–190 (2009).
96. Mawlawi, O. *et al.* Imaging human mesolimbic dopamine transmission with positron emission tomography: I. Accuracy and precision of D₂ receptor parameter measurements in ventral striatum. *J. Cereb. Blood Flow Metab.* **21**, 1034–1057 (2001).
97. Do we need an ethics of self-organizing tissue? *Nat. Methods* **12**, 895 (2015).
98. Lindgren, J. E., Ohlsson, A., Agurell, S., Hollister, L. & Gillespie, H. Clinical effects and plasma levels of Δ^9 -tetrahydrocannabinol (Δ^9 -THC) in heavy and light users of cannabis. *Psychopharmacology (Berl.)* **74**, 208–212 (1981).
99. Hardwick, S. K. L. *Home Office Cannabis Potency Study* (Home Office, 2008).
100. Hunault, C. C. *et al.* Disposition of smoked cannabis with high Δ^9 -tetrahydrocannabinol content: a kinetic model. *Toxicol. Appl. Pharmacol.* **246**, 148–153 (2010).
101. Banerjee, S. P., Snyder, S. H. & Mechoulam, R. Cannabinoids: influence on neurotransmitter uptake in rat brain synaptosomes. *J. Pharmacol. Exp. Ther.* **194**, 74–81 (1975).
102. Bloomfield, M. A., Morgan, C. J., Kapur, S., Curran, H. V. & Howes, O. D. The link between dopamine function and apathy in cannabis users: an [18 F]-DOPA PET imaging study. *Psychopharmacology (Berl.)* **231**, 2251–2259 (2014).
103. von Sydow, K., Lieb, R., Pfister, H., Hofler, M. & Wittchen, H. U. What predicts incident use of cannabis and progression to abuse and dependence? A 4-year prospective examination of risk factors in a community sample of adolescents and young adults. *Drug Alcohol Depend.* **68**, 49–64 (2002).
104. Carlini, E. A., Lindsey, C. J. & Tufik, S. Cannabis, catecholamines, rapid eye movement sleep and aggressive behaviour. *Br. J. Pharmacol.* **61**, 371–379 (1977).
105. MacLean, K. I. & Littleton, J. M. Environmental stress as a factor in the response of rat brain catecholamine metabolism to Δ^8 -tetrahydrocannabinol. *Eur. J. Pharmacol.* **41**, 171–182 (1977).
106. Lomax, P. Acute tolerance to the hypothermic effect of marijuana in the rat. *Res. Commun. Chem. Pathol. Pharmacol.* **2**, 159–167 (1971).
107. Andén, N. E. Dopamine turnover in the corpus striatum and the limbic system after treatment with neuroleptic and anti-acetylcholine drugs. *J. Pharm. Pharmacol.* **24**, 905–906 (1972).
108. Hattendorf, C., Hattendorf, M., Coper, H. & Fernandes, M. Interaction between Δ^9 -tetrahydrocannabinol and d-amphetamine. *Psychopharmacology (Berl.)* **54**, 177–182 (1977).
109. Williams, C. M., Rogers, P. J. & Kirkham, T. C. Hyperphagia in pre-fed rats following oral Δ^9 -THC. *Physiol. Behav.* **65**, 343–346 (1998).
110. Foltin, R. W., Brady, J. V. & Fischman, M. W. Behavioral analysis of marijuana effects on food intake in humans. *Pharmacol. Biochem. Behav.* **25**, 577–582 (1986).
111. Ungerstedt, U. Adipsia and aphagia after 6-hydroxydopamine induced degeneration of the nigro-striatal dopamine system. *Acta Physiol. Scand. Suppl.* **367**, 95–122 (1971).
112. Verty, A., McGregor, I. & Mallet, P. The dopamine receptor antagonist SCH 23390 attenuates feeding induced by Δ^9 -tetrahydrocannabinol. *Brain Res.* **1020**, 188–195 (2004).
113. Koch, M. *et al.* Hypothalamic POMC neurons promote cannabinoid-induced feeding. *Nature* **519**, 45–50 (2015).
114. Andrews, Z. B. *et al.* Ghrelin promotes and protects nigrostriatal dopamine function via a UCP2-dependent mitochondrial mechanism. *J. Neurosci.* **29**, 14057–14065 (2009).
115. Fergusson, D. M., Horwood, L. J. & Beutrais, A. L. Cannabis and educational achievement. *Addiction* **98**, 1681–1692 (2003).
- Longitudinal birth cohort study indicating that cannabis use in adolescents and early adults is associated with reduced educational achievement.**
116. Hooker, W. D. & Jones, R. T. Increased susceptibility to memory intrusions and the Stroop interference effect during acute marijuana intoxication. *Psychopharmacology (Berl.)* **91**, 20–24 (1987).
117. Ranganathan, M. & D'Souza, D. C. The acute effects of cannabinoids on memory in humans: a review. *Psychopharmacology (Berl.)* **188**, 425–444 (2006).
118. McGlothlin, W. H. W. L. The marijuana problem: an overview. *Am. J. Psychiatry* **125**, 126–134 (1968).
119. Levy, R. & Dubois, B. Apathy and the functional anatomy of the prefrontal cortex–basal ganglia circuits. *Cereb. Cortex* **16**, 916–928 (2005).
120. Goldman-Rakic, P. S. Regional and cellular fractionation of working memory. *Proc. Natl Acad. Sci. USA* **93**, 13473–13480 (1996).
121. Sawaguchi, T. & Goldman-Rakic, P. S. D₁ dopamine receptors in prefrontal cortex: involvement in working memory. *Science* **251**, 947–950 (1991).
122. Nava, F., Carta, G. & Gessa, G. Permissive role of dopamine D₂ receptors in the hypothermia induced by Δ^9 -tetrahydrocannabinol in rats. *Pharmacol. Biochem. Behav.* **66**, 183–187 (2000).
123. D'Souza, D. C. *et al.* Effects of haloperidol on the behavioral, subjective, cognitive, motor, and neuroendocrine effects of Δ^9 -tetrahydrocannabinol in humans. *Psychopharmacology (Berl.)* **198**, 587–603 (2008).
124. Tunbridge, E. M. *et al.* Genetic moderation of the effects of cannabis: catechol-O-methyltransferase (COMT) affects the impact of Δ^9 -tetrahydrocannabinol (THC) on working memory performance but not on the occurrence of psychotic experiences. *J. Psychopharmacol.* **29**, 1146–1151 (2015).
125. Paule, M. G. *et al.* Chronic marijuana smoke exposure in the rhesus monkey. II: Effects on progressive ratio and conditioned position responding. *J. Pharmacol. Exp. Ther.* **260**, 210–222 (1992).
126. Campbell, I. The amotivational syndrome and cannabis use with emphasis on the Canadian scene. *Ann. NY Acad. Sci.* **282**, 33–36 (1976).
127. Howes, O. D. *et al.* The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Arch. Gen. Psychiatry* **69**, 776–786 (2012).

128. Voruganti, L. N., Slomka, P., Zabel, P., Mattar, A. & Awad, A. G. Cannabis induced dopamine release: an *in-vivo* SPECT study. *Psychiatry Res.* **107**, 173–177 (2001).
129. Mizrahi, R. *et al.* Stress-induced dopamine response in subjects at clinical high risk for schizophrenia with and without concurrent cannabis use. *Neuropsychopharmacology* **39**, 1479–1489 (2014).
130. Thompson, J. L. *et al.* Striatal dopamine release in schizophrenia comorbid with substance dependence. *Mol. Psychiatry* **18**, 909–915 (2013).
131. Leweke, F. M. *et al.* Anandamide levels in cerebrospinal fluid of first-episode schizophrenic patients: Impact of cannabis use. *Schizophr. Res.* **94**, 29–36 (2007).
132. Leweke, F. M. *et al.* Cannabidiol enhances anandamide signaling and alleviates psychotic symptoms of schizophrenia. *Transl. Psychiatry* **2**, e94 (2012).
133. Kearn, C. S., Blake-Palmer, K., Daniel, E., Mackie, K. & Glass, M. Concurrent stimulation of cannabinoid CB₁ and dopamine D₂ receptors enhances heterodimer formation: a mechanism for receptor cross-talk? *Mol. Pharmacol.* **67**, 1697–1704 (2005).
134. Di Forti, M. *et al.* Confirmation that the AKT1 (rs2494732) genotype influences the risk of psychosis in cannabis users. *Biol. Psychiatry* **72**, 811–816 (2012).
135. Parsons, L. H. & Hurd, Y. L. Endocannabinoid signalling in reward and addiction. *Nat. Rev. Neurosci.* **16**, 579–594 (2015).
Impaired endocannabinoid signalling dysregulates synaptic plasticity and increases stress responsivity, negative emotional states and cravings that propel addiction.
136. Lecca, S., Melis, M., Luchicchi, A., Muntoni, A. L. & Pistis, M. Inhibitory inputs from rostromedial tegmental neurons regulate spontaneous activity of midbrain dopamine cells and their responses to drugs of abuse. *Neuropsychopharmacology* **37**, 1164–1176 (2012).
137. Marinelli, S. *et al.* N-arachidonoyl-dopamine tunes synaptic transmission onto dopaminergic neurons by activating both cannabinoid and vanilloid receptors. *Neuropsychopharmacology* **32**, 298–308 (2007).

Acknowledgements We thank V. M. Rajagopal and Nature for assistance with illustrations. This work was funded by a Medical Research Council (UK) Grant to O.D.H. (MC-A656-5QD30).

Author Contributions M.A.P.B. and O.D.H. conceptualized this review. M.A.P.B. and A.H.A. systematically reviewed the literature. All the authors contributed intellectually.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.H. (oliver.howes@csc.mrc.ac.uk).

Reviewer Information Nature thanks P. Fadda and R. Mechoulam for their contribution to the peer review of this work.

Forward-genetics analysis of sleep in randomly mutagenized mice

Hiromasa Funato^{1,2}, Chika Miyoshi^{1*}, Tomoyuki Fujiyama^{1*}, Takeshi Kanda^{1*}, Makito Sato^{1,3*}, Zhiqiang Wang¹, Jing Ma¹, Shin Nakane⁴, Jun Tomita⁴, Aya Ikkyu¹, Miyo Kakizaki¹, Noriko Hotta-Hirashima¹, Satomi Kanno¹, Haruna Komiya¹, Fuyuki Asano¹, Takato Honda¹, Staci J. Kim¹, Kanako Harano¹, Hiroki Muramoto¹, Toshiya Yonezawa¹, Seiya Mizuno⁵, Shinichi Miyazaki¹, Linzi Connor¹, Vivek Kumar^{6,7}, Ikuro Miura⁸, Tomohiro Suzuki⁸, Atsushi Watanabe⁹, Manabu Abe¹⁰, Fumihiko Sugiyama⁵, Satoru Takahashi⁵, Kenji Sakimura¹⁰, Yu Hayashi^{1,11}, Qinghua Liu^{1,12}, Kazuhiko Kume⁴, Shigeharu Wakana⁸, Joseph S. Takahashi^{1,6,13} & Masashi Yanagisawa^{1,3,13,14}

Sleep is conserved from invertebrates to vertebrates, and is tightly regulated in a homeostatic manner. The molecular and cellular mechanisms that determine the amount of rapid eye movement sleep (REMS) and non-REMS (NREMS) remain unknown. Here we identify two dominant mutations that affect sleep and wakefulness by using an electroencephalogram/electromyogram-based screen of randomly mutagenized mice. A splicing mutation in the *Sik3* protein kinase gene causes a profound decrease in total wake time, owing to an increase in inherent sleep need. Sleep deprivation affects phosphorylation of regulatory sites on the kinase, suggesting a role for SIK3 in the homeostatic regulation of sleep amount. *Sik3* orthologues also regulate sleep in fruitflies and roundworms. A missense, gain-of-function mutation in the sodium leak channel *NALCN* reduces the total amount and episode duration of REMS, apparently by increasing the excitability of REMS-inhibiting neurons. Our results substantiate the use of a forward-genetics approach for studying sleep behaviours in mice, and demonstrate the role of SIK3 and *NALCN* in regulating the amount of NREMS and REMS, respectively.

Sleep is an animal behaviour ubiquitously conserved from vertebrates to invertebrates, including flies and nematodes^{1–3}, and is tightly regulated in a homeostatic manner. Sleep in mammals exhibits the cycles of REMS and NREMS that are defined by characteristic activity of electroencephalogram (EEG) and electromyogram (EMG). Time spent in sleep is determined by a homeostatic sleep need, a driving force for sleep/wakefulness switching, which increases during wakefulness and dissipates during sleep^{4,5}. The spectral power in the delta-range frequency (1–4 Hz) of EEG during NREMS has been regarded as one of best markers for the current level of sleep need. Conversely, the level of arousal is positively correlated with sleep latency, which can be regulated independently of sleep need⁶, reflecting the overall activity of wake-promoting neurons. Traditional approaches to locate the neural circuits regulating sleep and wakefulness behaviour included local ablation of brain regions^{7–9}. Recent advances in optogenetic and chemogenetic research have directly demonstrated that switching between sleep and wake states is executed by subsets of neurons in the basal forebrain¹⁰, lateral hypothalamus^{11,12} and locus coeruleus¹³, and that switching between NREMS and REMS is executed by a neural network in the pons and medulla^{14,15}. Despite the accumulating information about executive neural circuitries regulating sleep and wake states, the molecular and cellular mechanisms that determine the propensity of switching between wakefulness, REMS and NREMS remain unknown. To tackle this problem, we have used a phenotype-driven forward-genetics approach that is free from specific working hypotheses¹⁶. Previously, a series of forward-genetics studies using flies

and mice successfully uncovered the molecular network of the core clock genes regulating circadian behaviours^{17–19}. Sleep-regulating genes were also discovered through the screening of mutagenized flies^{1,2}. However, genetic studies for sleep using mice has been challenging because of the effective compensation and redundancy in the regulation of sleep and wakefulness, and the need for EEG and EMG monitoring of the staging of wakefulness, NREMS and REMS.

Sik3 splice mutation increases NREMS

We induced random point mutations into C57BL/6J (B6J) male mice (G_0) by ethylnitrosourea (ENU) and screened more than 8,000 heterozygous B6J \times C57BL/6N (B6N) F_1 mice for dominant sleep and wakefulness abnormalities through EEG/EMG-based sleep staging (Extended Data Fig. 1a). B6N was chosen as a counter strain because its sleep and wakefulness parameters are highly similar to B6J (Extended Data Fig. 1b), and the entire list of single nucleotide polymorphisms has recently become available²⁰.

Through our screening, we established a mutant pedigree, which we termed *Sleepy* (*Slp*), with a markedly prolonged sleep time. Five founders of the *Sleepy* mutant pedigree were born by *in vitro* fertilization using sperm from the same ENU-treated G_0 male, and showed a daily wake time (524 ± 19.7 min; mean \pm s.d.) that was shorter than the mean of all mice screened by >3 standard deviations (Extended Data Fig. 1c). The *Sleepy* pedigree showed clear dominant inheritance of reduced wake time (Fig. 1a). Linkage analysis in the B6J \times B6N N_2 generation of five *Sleepy* pedigrees (B021–B025) produced a single log

¹International Institute for Integrative Sleep Medicine (WPI-IIS), University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan. ²Department of Anatomy, Faculty of Medicine, Toho University, Ota-ku, Tokyo 143-8540, Japan. ³Department of Molecular Genetics, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁴Department of Neuropharmacology, Graduate School of Pharmaceutical Sciences, Nagoya City University, Nagoya, Aichi 467-8603, Japan. ⁵Laboratory Animal Resource Center, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan. ⁶Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁷The Jackson Laboratory, Bar Harbor, Maine 04609, USA. ⁸Technology and Development Team for Mouse Phenotype Analysis, RIKEN Bioresource Center, Tsukuba, Ibaraki 305-0074, Japan. ⁹Laboratory of Research Advancement, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan. ¹⁰Department of Cellular Neurobiology, Brain Research Institute, Niigata University, Niigata 951-8585, Japan. ¹¹PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. ¹²Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ¹³Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ¹⁴Life Science Center, Tsukuba Advanced Research Alliance, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan.

*These authors contributed equally to this work.

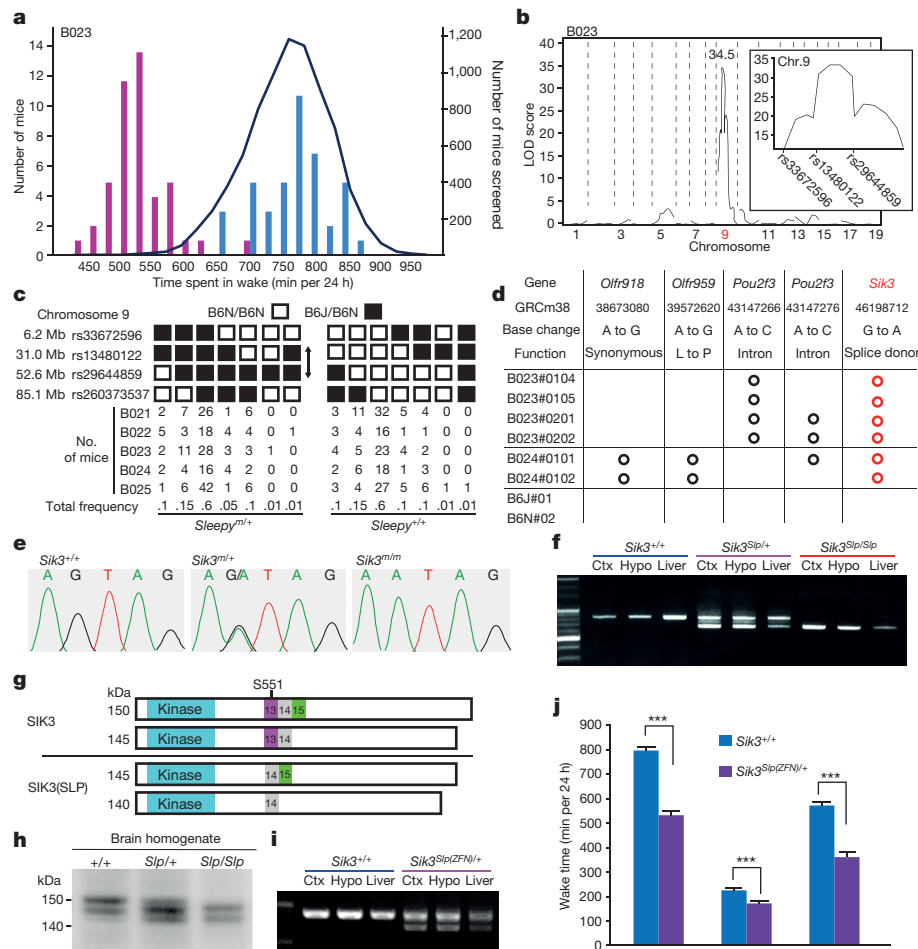


Figure 1 | Identification of *Sik3* splicing mutation responsible for reduced total wake time. **a**, Wake-time distribution of B023 N₂ littermates (bars) and all mice screened (curve). Blue and purple bars indicates retrospectively genotyped *Sik3*^{+/+} and *Sik3*^{Slp/+} mice, respectively. **b**, Quantitative trait locus (QTL) analysis of B023 pedigree ($n = 93$) for total wake time. Inset shows LOD score peak between rs13480122 and rs29644859. **c**, Haplotype analysis of *Sleepy* mutant pedigrees, B021–B025, in terms of the presence of *Sleepy* phenotypes. **d**, Exome sequencing results from *Sleepy* mutant mice of B023 and B024 pedigrees together with wild-type mice within the region of the LOD score peak. **e**, Direct

odds ratio (LOD) score peak on chromosome 9 (Fig. 1b and Extended Data Fig. 2a), between rs13480122 (chr9: 31156626) and rs29644859 (chr9: 52785119) (Fig. 1b, c). Whole-exome sequencing of *Sleepy* mutants identified a heterozygous single nucleotide substitution at the splice donor site (chr9: 46198712) for intron 13 of the *Sik3* gene (Fig. 1d, e). The mutation predicted an abnormal skipping of exon 13 (*Sik3*^{Slp/+}), which was confirmed by reverse transcription PCR (RT-PCR) and sequencing of the *Sik3* mRNA (Fig. 1f, Supplementary Fig. 1 and Extended Data Fig. 2b).

SIK3 is a protein kinase that is broadly expressed in neurons of the cerebral cortex, thalamus, hypothalamus and brainstem (Extended Data Fig. 2c, d), and belongs to the AMP-activated protein kinase (AMPK) family. SIK3 has a serine-threonine kinase domain at the N terminus and a protein kinase A (PKA) recognition site (Ser551) in the middle portion²¹ (Fig. 1g). The skipping of exon 13 results in an in-frame deletion of 52 amino acids, encompassing the PKA site (Fig. 1g, Extended Data Fig. 2e and Supplementary Fig. 1). Immunoblotting of brain homogenates from *Sik3*^{Slp/+} and *Sik3*^{Slp/Slp} mice using anti-SIK3 antibody detected smaller SIK3 proteins as predicted (Fig. 1g, h and Supplementary Fig. 1). The nature of this mutation, taken together with our results on SIK3 orthologues in *Drosophila melanogaster*

sequencing of the *Sik3* gene. **f**, RT-PCR of *Sik3* mRNA produced smaller bands specific to *Sik3*^{Slp} mice. Ctx, cerebral cortex; Hypo, hypothalamus. **g**, Structures of wild-type and mutant SIK3 proteins. **h**, Immunoblotting of brain homogenates showing wild-type (+/+) and mutant SIK3 (*Slp*/+ and *Slp*/*Slp*) protein variants. **i**, RT-PCR of brain mRNA from ZFN-based *Sik3*^{Slp/+} mice showing smaller bands lacking exon 13, in addition to larger bands containing the exon 13. **j**, *Sik3*^{Slp(ZFN)/+} mice ($n = 15$) showed a shorter total wake time than wild-type littermates ($n = 14$). *** $P < 0.001$, two-way analysis of variance (ANOVA) followed by Tukey's test. Data are mean \pm s.e.m.

and *Caenorhabditis elegans* (see below), suggests that *Sik3*^{Slp} is a gain-of-function allele.

To confirm genetically that the *Sik3* splice mutation is the sole cause of the long-sleep phenotype, we introduced the *Sik3* exon 13-skipping allele into wild-type mice by using either conventional knock-in in embryonic stem (ES) cells or zinc-finger nuclease (ZFN) technology (Fig. 1i, Extended Data Fig. 3b and Supplementary Fig. 1). As expected, these mouse lines exhibited markedly reduced wake time (Fig. 1j, Extended Data Fig. 3c), similar to the original *Sleepy* pedigree. Thus, the lack of exon 13 of the *Sik3* gene causes the *Sleepy* phenotype.

Increased sleep need in *Sleepy* mutants

No overt abnormality in the sleep EEG/EMG was seen in the *Sleepy* mutants (Extended Data Fig. 4a). Detailed examination of sleep and wakefulness behaviour of *Sleepy* mutants showed that, although *Sik3*^{Slp/+} mice exhibit reduced wake time and increased NREMS time both in the light and dark phases, the hypersomnia phenotype is more pronounced in the dark phase (Fig. 2a, b), possibly owing to a ceiling effect in the light phase. Homozygous *Sik3*^{Slp/Slp} mice had an even shorter total wake time and longer NREMS time than *Sik3*^{Slp/+} mice (Fig. 2a, b and Extended Data Fig. 4b–d), showing an allele dosage

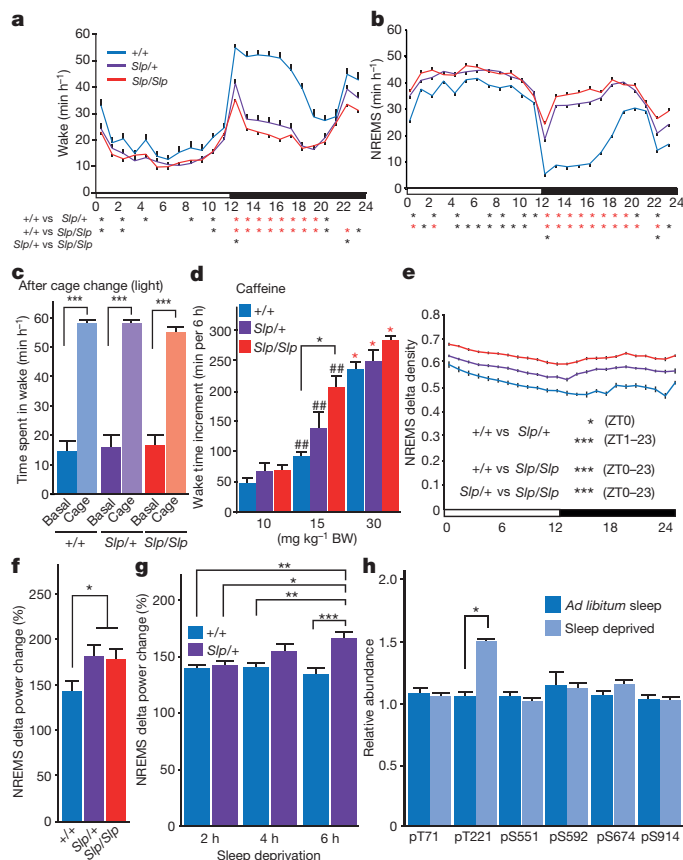


Figure 2 | Increased sleep need and normal wake-promoting response of *Sik3* mutant mice. **a, b**, Circadian variation in wakefulness (**a**) and NREMS (**b**) in *Sik3*^{+/+} (*n* = 22), *Sik3*^{Slp/+} (*n* = 32) and *Sik3*^{Slp/Slp} (*n* = 31) mice. **P* < 0.05 (black); ***P* < 0.001 (red), one-way repeated measures ANOVA followed by Tukey's test. **c**, Time spent in wakefulness from zeitgeber time (ZT)4 to ZT5 after cage change at ZT5 of *Sik3*^{+/+} (*n* = 6), *Sik3*^{Slp/+} (*n* = 9) and *Sik3*^{Slp/Slp} (*n* = 6) mice. ****P* < 0.001, one-way repeated measures ANOVA followed by Tukey's test. **d**, Wake time for 6 h after caffeine injection at ZT0 in *Sik3*^{+/+} (*n* = 6), *Sik3*^{Slp/+} (*n* = 6) and *Sik3*^{Slp/Slp} (*n* = 6) mice. **P* < 0.05 (black); ##*P* < 0.01 versus 10 mg kg⁻¹ body weight (BW); **P* < 0.01 (red) versus 15 mg kg⁻¹ BW, two-way ANOVA followed by Tukey's test. **e**, NREMS delta density of *Sik3* mutant mice (*Sik3*^{+/+}, *n* = 22; *Sik3*^{Slp/+}, *n* = 32; *Sik3*^{Slp/Slp}, *n* = 31) across the light-dark cycle. **P* < 0.05; ****P* < 0.001, one-way repeated ANOVA followed by Tukey's test. **f**, Increase in NREMS delta power of *Sik3*^{+/+} (*n* = 7), *Sik3*^{Slp/+} (*n* = 7) and *Sik3*^{Slp/Slp} (*n* = 10) mice after 6-h sleep deprivation. **P* < 0.05, one-way ANOVA followed by Tukey's test. **g**, Increase in NREMS delta power after 2, 4 and 6 h of sleep deprivation of *Sik3*^{+/+} (*n* = 11) and *Sik3*^{Slp/+} (*n* = 11) mice relative to NREMS delta power of the same ZT during basal sleep. **P* < 0.05; ***P* < 0.01; ****P* < 0.001, two-way ANOVA followed by Tukey's test. **h**, Phosphorylation of Flag-SIK3 of *Flag-Sik3*^{+/+} brains with or without 4-h sleep deprivation. **P* < 0.05, two-way ANOVA followed by Tukey's test. Data are mean ± s.e.m.

effect. Total REMS time was similar among genotype groups (Extended Data Fig. 4e). *Sik3*^{Slp/+} and *Sik3*^{Slp/Slp} mice had a higher NREMS/total sleep ratio (Extended Data Fig. 4f) and a lower REMS/total sleep ratio (Extended Data Fig. 4g), indicating a NREMS-specific change in the *Sleepy* mutant mice. Longer time spent in REMS during the dark phase seems to be secondary to the increased sleep amount and be compensated with shorter REMS time during the light phase (Extended Data Fig. 4h). Female *Sik3*^{Slp/+} and *Sik3*^{Slp/Slp} mice also exhibited a similar phenotype (Extended Data Fig. 4i).

We next examined whether the short wake time of *Sleepy* mice is due to a defective wake-promoting response to behavioural or pharmacological stimuli. A new cage environment strongly mobilizes the wake-promoting system in mice⁶. Both *Sik3*^{Slp/+} and *Sik3*^{Slp/Slp}

mice exhibited prominent wake responses to cage change, similar to wild-type littermates during the light phase (Fig. 2c) and dark phase (Extended Data Fig. 5a). *Sik3*^{Slp/+} and *Sik3*^{Slp/Slp} mice also showed increased wake time in response to the administration of caffeine (Fig. 2d) or modafinil (Extended Data Fig. 5b), to similar or even slightly higher degrees than *Sik3*^{+/+} littermates. We then examined the behavioural circadian rhythm of the *Sleepy* mutants. The circadian period length as assessed by wheel-running behaviour under constant darkness was similar among genotype groups (Extended Data Fig. 5c). *Sleepy* mutant mice showed a robust reduction in wake time also under constant darkness, similar to light–dark conditions (Extended Data Fig. 5d).

Given the apparently normal wake-promoting and circadian systems, we then proposed that the *Sleepy* mutants have an inherently higher sleep need. Indeed, *Sleepy* mutant mice exhibited a higher density of slow-wave activity during NREMS in an allele-dosage dependent fashion (Fig. 2e), suggesting that the baseline sleep need is increased in the mutants. Lower-frequency power in EEG was also increased during wakefulness in *Sleepy* mutant mice (Extended Data Fig. 5e), which may be associated with local cortical synchronization owing to increased sleep need²². Furthermore, 6 h of sleep deprivation from the onset of the light phase increased the NREMS delta power of *Sik3*^{Slp} mice to a larger extent than *Sik3*^{+/+} mice (Fig. 2f). To examine dose-dependent effects, we then conducted 2, 4 and 6 h of sleep deprivation. Under our conditions, *Sik3*^{+/+} mice showed only a slight increase after 2 h of sleep deprivation, whereas *Sik3*^{Slp/+} mice exhibited a marked dose-dependent increase in the delta power (Fig. 2g and Extended Data Fig. 5f), demonstrating their exaggerated response to sleep deprivation.

If SIK3 constitutes a part of the enigmatic molecular pathway determining the level of homeostatic sleep need, then the protein should be modulated by sleep deprivation. Because the function of SIK3 is regulated by its own phosphorylation²³, we examined its phosphorylation status in the brain from mice in which a Flag-tag was inserted into the N terminus of SIK3 using CRISPR/Cas9 technology (Extended Data Fig. 6a). Immunoblotting and immunoprecipitation with anti-SIK3 and anti-Flag antibodies detected Flag-SIK3 protein in the brain as expected (Extended Data Fig. 6b–d and Supplementary Fig. 1). To examine whether sleep deprivation affects the phosphorylation status of SIK3, brains of *Flag-Sik3* mice were obtained after sleep deprivation. Quantitative phosphoproteomic analysis of immunopurified Flag-tagged proteins showed that sleep deprivation specifically increases the phosphorylation of residue Thr221, which is closely associated with the kinase activity of SIK3 (ref. 23; Fig. 2h). In the same way, we introduced a Flag sequence into the *Sik3*^{Slp} allele, and obtained mice expressing mutant Flag-SIK3(SLP) protein (Extended Data Fig. 6b, c and Supplementary Fig. 1). In addition to the expected decrease in the phosphorylation of Ser551, Flag-SIK3(SLP) in the brains showed a decrease in the phosphorylation of Thr469, another PKA site²⁴, and an increase in the phosphorylation of Ser914 compared with *Flag-Sik3* mice under ad libitum sleep during the light phase (Extended Data Fig. 5g). After sleep deprivation, these phosphorylation changes were sustained and a significant increase in the phosphorylation of Ser674, another PKA site²⁴, was recognized (Extended Data Fig. 5g). Thus, the lack of the Ser551-containing region disturbs the phosphorylation status of other PKA sites in a complex fashion, suggesting a mechanistic link with the increased sleep need of *Sleepy* mutant mice.

Conserved role of SIK3 in invertebrates

The exon 13-encoded region of *Sik3* is highly conserved among vertebrate animals (Fig. 3a). Notably, there is a PKA recognition site within the exon 13-encoded region, which is conserved even in *Sik3* orthologues of *Drosophila* and *C. elegans* (Fig. 3a and Extended Data Fig. 7). To examine the role of *Sik3* orthologues in sleep-like behaviours in *Drosophila*, we expressed a phosphorylation-defective SIK3 (Ser563Ala; serine residue equivalent to Ser551 in mouse SIK3) in neuronal cells in an RU486-inducible fashion. Daily sleep time of adult flies is increased

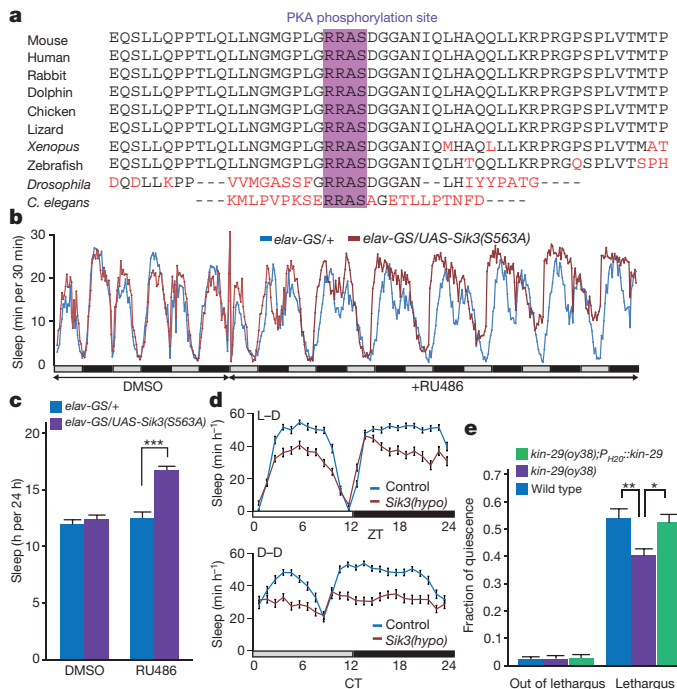


Figure 3 | Role of *Sik3* orthologues in invertebrate sleep-like behaviours.

a, The phylogenetic conservation of the exon 13-encoded region of *Sik3*. **b**, **c**, Sleep time before and after induction of *Sik3*(Ser563Ala) by RU486 under constant darkness (32 per group). **d**, Sleep time of control and *Sik3* hypomorphic mutant (*Sik3(hypo)*) in 12-h light:12-h dark (L-D) conditions (top, 16 per group, $P < 0.001$) and in constant darkness (D-D) (bottom, 16 per group, $P < 0.001$); one-way repeated measures ANOVA. CT, circadian time. **e**, *Sik3*-null mutant worms, *kin-29(oy38)* ($n = 15$), exhibited reduced quiescence during lethargus compared with wild-type worms ($n = 10$). *kin-29(oy38);P_{H2O}::kin-29* worms ($n = 9$), in which wild-type *kin-29* was expressed in neuronal cells, restored normal quiescence during lethargus. The fractions of quiescence out of lethargus were similar ($P = 0.98$). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$, one-way repeated measures ANOVA followed by Tukey's test. Data are mean \pm s.e.m.

after induction of *Sik3*(Ser563Ala) (Fig. 3b, c). Conversely, flies with a hypomorphic *Sik3* mutation showed reduced sleep time in the light–dark cycle and in constant darkness (Fig. 3d). Furthermore, a null mutation of *kin-29*, the *C. elegans* orthologue of *Sik3*, reduced the fraction of quiescence during L4-adult lethargus, a sleep-like state in *C. elegans*³ (Fig. 3e). Pan-neuronal expression of *kin-29* rescued this phenotype. Thus, orthologues of *Sik3* also seem to be involved in the regulation of sleep amount in fruitflies and nematodes.

Nalcn mutation reduces REMS

The EEG/EMG-based dominant screening of ENU-mutagenized mice also yielded a mutant pedigree with a REMS abnormality, which we termed *Dreamless* (*Drl*). The founder of the pedigree showed a short REMS episode duration (Fig. 4a), that was heritable in the offspring (Extended Data Fig. 8a). Linkage analysis in the B6J \times B6N N₂ generation showed a single LOD score peak near rs31233932 (chr14: 124108797) on chromosome 14 (Fig. 4b and Extended Data Fig. 8b). Whole-exome sequencing combined with direct sequencing of candidate genes identified a heterozygous missense mutation in the *Nalcn* gene (chr14: 123515403) as the only functionally relevant mutation within the mapped region in *Dreamless* mutants (Fig. 4c and Extended Data Fig. 8c). We then confirmed the causal relationship of the *Nalcn* gene mutation to the REMS phenotype by introducing the same nucleotide substitution in wild-type mice using the CRISPR/Cas9 system. CRISPR *Nalcn*^{Drl/+} mice showed a short REMS episode duration, similar to the original *Dreamless* pedigree (Fig. 4d–g).

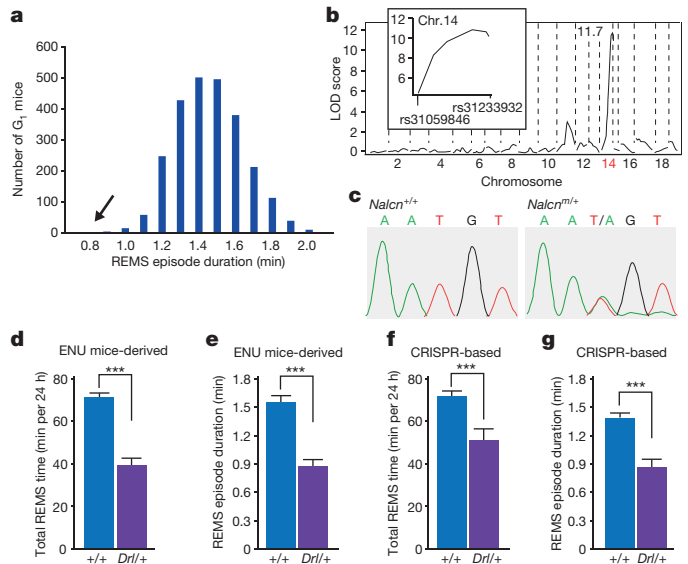


Figure 4 | Missense mutation in *Nalcn* gene reduces REMS time and episode duration. **a**, Histogram of REMS episode duration of G₁ mice screened (mean \pm s.d. = 1.41 ± 0.19 min). Arrow indicates the founder of *Dreamless* mutant pedigree. **b**, QTL analysis of *Dreamless* mutant pedigree ($n = 56$) for REMS episode duration. Inset shows LOD score peak near rs31233932. **c**, Direct sequencing of the *Nalcn* gene. **d**, **e**, Total REMS time (**d**) and REMS episode duration (**e**) of *Nalcn*^{Drl/+} mice ($n = 29$) and *Nalcn*^{+/+} mice ($n = 25$) of the *Dreamless* mutant pedigree. **f**, **g**, Total REMS time (**f**) and REMS episode duration (**g**) of *Nalcn*^{Drl/+} mice ($n = 11$) and *Nalcn*^{+/+} mice ($n = 17$) produced by CRISPR/Cas9 technology. $***P < 0.001$, two-tailed Student's *t*-test. Data are mean \pm s.e.m.

Dreamless mutants showed theta (6–9 Hz)-dominant EEG and appropriate muscle atonia during REMS, and did not show any overt abnormality in the EEG/EMG (Extended Data Fig. 9a). *Nalcn*^{Drl/+} mice showed a reduced total REMS time owing to a short average REMS episode duration (Fig. 4d–g and Extended Data Fig. 9b, c). In other words, *Nalcn*^{Drl/+} mice failed to maintain the REMS state properly, resulting in highly unstable REMS. The total time spent in the wake and NREMS states of *Nalcn*^{Drl/+} mice was similar to that of *Nalcn*^{+/+} mice (Extended Data Fig. 9d). *Nalcn*^{Drl/+} mice had normal circadian period lengths but showed a greatly reduced amplitude of behavioural circadian rhythms under constant darkness (Extended Data Fig. 9e, f), consistent with the recently reported involvement of NALCN in the circadian regulation of neuronal excitability in the suprachiasmatic nucleus²⁵. A marked reduction in the REMS time of *Nalcn*^{Drl/+} mice was also observed under constant darkness (Extended Data Fig. 9g). Spectral analysis of EEG showed a decrease in the theta-range power during NREMS and REMS and an increase in low-frequency power during the wake state and REMS in *Nalcn*^{Drl/+} mice (Extended Data Fig. 9h), suggesting the possibility that NALCN may regulate various oscillations in the brain.

NALCN is a voltage-independent, non-selective cation channel with a proposed role in the control of neuronal excitability²⁶. It is highly expressed in several brainstem nuclei involved in REMS regulation, such as the ventrolateral periaqueductal grey, deep mesencephalic nucleus (DpMe) and sublaterodorsal nucleus^{7,9,14,15} (Extended Data Fig. 10a–c). The mutation results in an Asn315Lys substitution in helix S6 of domain I (Fig. 5a), which is conserved among vertebrates and invertebrates (Fig. 5b).

To examine whether the mutation changes the electrophysiological properties of NALCN, we made patch-clamp recordings from HEK293 cells cotransfected with NALCN or the NALCN(DRL) mutant, together with UNC80 and SRC(Tyr529Phe), which allow constitutive activation of NALCN^{27,28}. Both NALCN and NALCN(DRL) showed linear current–voltage relationships (Fig. 5c, d and Extended Data Fig. 10d, e),

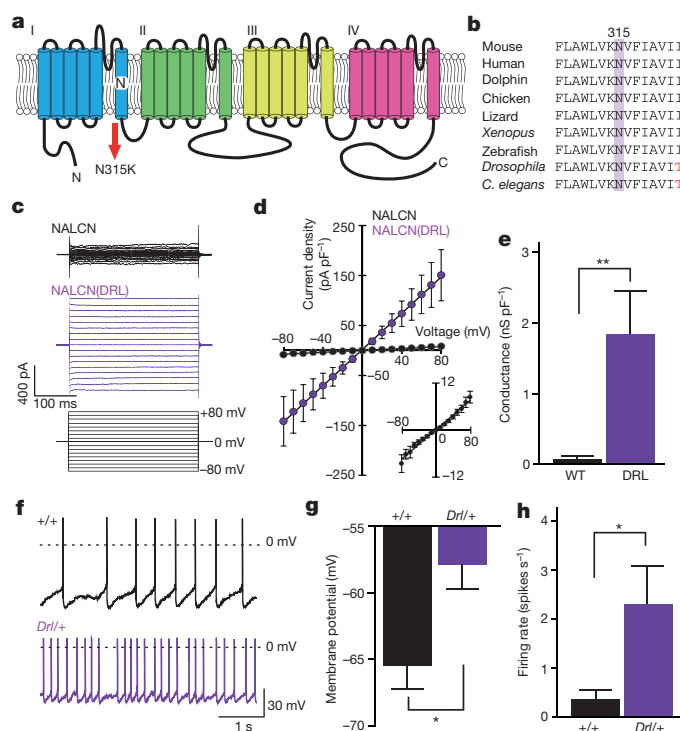


Figure 5 | Dreamless mutation in *Nalcn* gene increases excitability of neurons in the ‘REM-off’ area. **a**, Schematic structure of NALCN protein. **b**, Phylogenetic conservation of Asn315 residue in NALCN. **c**, Representative traces of membrane currents in response to 300-ms step pulses ranging from -80 mV to $+80$ mV in 10 -mV increments (holding potential $V_h = 0$ mV, bottom) recorded from HEK293T cells transfected with wild-type NALCN (top) or mutant NALCN(DRL) (middle). **d**, Mean current–voltage (I – V) curves in wild-type NALCN ($n = 5$, black circles and bottom right) or NALCN(DRL) ($n = 7$, purple circles). **e**, The conductance of NALCN(DRL)-transfected cells was larger than that of wild-type NALCN-transfected cells (NALCN, 0.09 ± 0.02 nS pF $^{-1}$, $n = 5$; NALCN(DRL), 1.81 ± 0.62 nS pF $^{-1}$, $n = 7$). $^{**}P < 0.01$, Mann–Whitney U test. **f**, Representative trace of membrane potentials of DpMe neurons in *Nalcn* $^{+/+}$ (top) and *Nalcn* $^{Drl/+}$ (bottom) mice. Dashed lines indicate 0 mV level. **g**, **h**, Mean membrane potentials (**g**) and spontaneous firing rates (**h**) of DpMe neurons (*Nalcn* $^{+/+}$, $n = 33$; *Nalcn* $^{Drl/+}$, $n = 31$). $^{*}P < 0.05$, Mann–Whitney U test. Data are mean \pm s.e.m.

with similar equilibrium potentials (NALCN, -2.1 ± 1.6 mV, $n = 7$; NALCN(DRL), -1.2 ± 0.7 mV, $n = 5$, $P = 0.94$, Mann–Whitney U test). However, the ionic conductance of NALCN(DRL) was much larger than that of NALCN (Fig. 5e). Similarly, the charge transfer of inward and outward currents in NALCN(DRL)-transfected cells was markedly higher than in NALCN-transfected cells (Extended Data Fig. 10f), raising the possibility that NALCN(DRL) may increase the intrinsic excitability of REMS-regulatory neurons. To test this possibility, whole-cell current-clamp recordings were made from neurons in the DpMe (Extended Data Fig. 10a), which contains ‘REM-off’ cells^{15,29,30}, using brain slices from *Nalcn* $^{Drl/+}$ and *Nalcn* $^{+/+}$ mice. The DpMe neurons in *Nalcn* $^{Drl/+}$ slices exhibited depolarization (*Nalcn* $^{+/+}$, -65.6 ± 1.7 mV, $n = 33$ cells; *Nalcn* $^{Drl/+}$, -57.8 ± 2.0 mV, $n = 31$ cells; $P = 0.01$, Mann–Whitney U test) and higher spontaneous firing rates (*Nalcn* $^{+/+}$, 0.4 ± 2.3 spikes s $^{-1}$, $n = 33$ cells; *Nalcn* $^{Drl/+}$, 2.3 ± 0.8 spikes s $^{-1}$, $n = 31$ cells; $P = 0.03$, Mann–Whitney U test) compared with *Nalcn* $^{+/+}$ slices (Fig. 5f–h).

Discussion

Our EEG/EMG-based forward-genetics screen in mice has identified distinct sleep phenotypes and mutated genes. Yet our study also illustrates a conserved role of SIK3 in the sleep behaviour of vertebrates and invertebrates. We think that the *Sleepy* mutation in SIK3 increases

the animal’s intrinsic sleep need, because *Sleepy* mutant mice exhibit (1) a higher density of slow-wave activity, a reliable index of homeostatic sleep need; (2) a larger increase in NREMS delta power after sleep deprivation; and (3) a normal waking response to behavioural or pharmacological arousal stimuli. Furthermore, we found that the functionally relevant phosphorylation status of SIK3 is modulated by sleep deprivation, which induces an acute increase of sleep need. Taken together with the finding that the *Sleepy* mutation markedly increases the baseline NREMS amounts in an allele dosage-dependent fashion, we propose that SIK3 functions in the intracellular signalling pathway that dictates sleep need and homeostatically regulates the daily amount of sleep.

We propose that NALCN works in the neuronal groups regulating REMS^{9,14,15} for the maintenance and termination of REMS episodes. The narrow abdomen mutant, carrying a loss-of-function mutation in the *Drosophila* orthologue of NALCN, exhibits increased sensitivity to anaesthesia³¹ and abnormal circadian behaviour³² in part through a blunted circadian change in neuronal excitability²⁵. The crucial role of the narrow abdomen in enhancing bi-stability between wakefulness and the anaesthetized state, as well as between wakefulness and sleep³³, is consistent with de-stabilized REMS episodes of the *Dreamless* mutant mice. Thus, like the case of SIK3 above, NALCN orthologues also appear to have conceptually analogous roles in regulating sleep-related behaviours in both mice and fruitflies. Our results argue for the utility of unbiased forward-genetics screens for the discovery of genes, alleles and pathways regulating sleep in mammals.

Note added in proof: A related review article describing the neural circuitries controlling sleep and wake states can be found in ref. 34.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 April; accepted 6 October 2016.

Published online 2 November 2016.

- Cirelli, C. *et al.* Reduced sleep in *Drosophila* *Shaker* mutants. *Nature* **434**, 1087–1092 (2005).
- Koh, K. *et al.* Identification of SLEEPLESS, a sleep-promoting factor. *Science* **321**, 372–376 (2008).
- Raizen, D. M. *et al.* Lethargus is a *Caenorhabditis elegans* sleep-like state. *Nature* **451**, 569–572 (2008).
- Daan, S., Beersma, D. G. & Borbély, A. A. Timing of human sleep: recovery process gated by a circadian pacemaker. *Am. J. Physiol.* **246**, R161–R183 (1984).
- Franken, P., Chollet, D. & Tafti, M. The homeostatic regulation of sleep need is under genetic control. *J. Neurosci.* **21**, 2610–2621 (2001).
- Suzuki, A., Sinton, C. M., Greene, R. W. & Yanagisawa, M. Behavioral and biochemical dissociation of arousal and homeostatic sleep need influenced by prior wakeful experience in mice. *Proc. Natl Acad. Sci. USA* **110**, 10288–10293 (2013).
- Lu, J., Sherman, D., Devor, M. & Saper, C. B. A putative flip-flop switch for control of REM sleep. *Nature* **441**, 589–594 (2006).
- Saper, C. B., Scammell, T. E. & Lu, J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* **437**, 1257–1263 (2005).
- Luppi, P. H. *et al.* The neuronal network responsible for paradoxical sleep and its dysfunctions causing narcolepsy and rapid eye movement (REM) behavior disorder. *Sleep Med. Rev.* **15**, 153–163 (2011).
- Xu, M. *et al.* Basal forebrain circuit for sleep–wake control. *Nat. Neurosci.* **18**, 1641–1647 (2015).
- Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. & de Lecea, L. Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* **450**, 420–424 (2007).
- Herrera, C. G. *et al.* Hypothalamic feedforward inhibition of thalamocortical network controls arousal and consciousness. *Nat. Neurosci.* **19**, 290–298 (2016).
- Carter, M. E. *et al.* Tuning arousal with optogenetic modulation of locus coeruleus neurons. *Nat. Neurosci.* **13**, 1526–1533 (2010).
- Weber, F. *et al.* Control of REM sleep by ventral medulla GABAergic neurons. *Nature* **526**, 435–438 (2015).
- Hayashi, Y. *et al.* Cells of a common developmental origin regulate REM/non-REM sleep and wakefulness in mice. *Science* **350**, 957–961 (2015).
- Takahashi, J. S., Shimomura, K. & Kumar, V. Searching for genes underlying behavior: lessons from circadian rhythms. *Science* **322**, 909–912 (2008).

17. Citri, Y. *et al.* A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. *Nature* **326**, 42–47 (1987).
18. King, D. P. *et al.* Positional cloning of the mouse circadian *Clock* gene. *Cell* **89**, 641–653 (1997).
19. Allada, R., Emery, P., Takahashi, J. S. & Rosbash, M. Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**, 1091–1119 (2001).
20. Kumar, V. *et al.* C57BL/6N mutation in cytoplasmic FMRP interacting protein 2 regulates cocaine response. *Science* **342**, 1508–1512 (2013).
21. Takemori, H. & Okamoto, M. Regulation of CREB-mediated gene expression by salt inducible kinase. *J. Steroid Biochem. Mol. Biol.* **108**, 287–291 (2008).
22. Vyazovskiy, V. V. *et al.* Local sleep in awake rats. *Nature* **472**, 443–447 (2011).
23. Katoh, Y. *et al.* Silencing the constitutive active transcription factor CREB by the LKB1-SIK signaling cascade. *FEBS J.* **273**, 2730–2748 (2006).
24. Berggreen, C., Henriksson, E., Jones, H. A., Morrice, N. & Göransson, O. cAMP-elevation mediated by β -adrenergic stimulation inhibits salt-inducible kinase (SIK) 3 activity in adipocytes. *Cell. Signal.* **24**, 1863–1871 (2012).
25. Flourakis, M. *et al.* A conserved bicycle model for circadian clock control of membrane excitability. *Cell* **162**, 836–848 (2015).
26. Ren, D. Sodium leak channels in neuronal excitability and rhythmic behaviors. *Neuron* **72**, 899–911 (2011).
27. Lu, B. *et al.* Peptide neurotransmitters activate a cation channel complex of NALCN and UNC-80. *Nature* **457**, 741–744 (2009).
28. Lu, B. *et al.* Extracellular calcium controls background current and neuronal excitability via an UNC79-UNC80-NALCN cation channel complex. *Neuron* **68**, 488–499 (2010).
29. Crochet, S., Onoe, H. & Sakai, K. A potent non-monoaminergic paradoxical sleep inhibitory system: a reverse microdialysis and single-unit recording study. *Eur. J. Neurosci.* **24**, 1404–1412 (2006).
30. Sapin, E. *et al.* Localization of the brainstem GABAergic neurons controlling paradoxical (REM) sleep. *PLoS One* **4**, e4272 (2009).
31. Krishnan, K. S. & Nash, H. A. A genetic study of the anesthetic response: mutants of *Drosophila melanogaster* altered in sensitivity to halothane. *Proc. Natl Acad. Sci. USA* **87**, 8632–8636 (1990).
32. Lear, B. C. *et al.* The ion channel narrow abdomen is critical for neural output of the *Drosophila* circadian pacemaker. *Neuron* **48**, 965–976 (2005).
33. Joiner, W. J. *et al.* Genetic and anatomical basis of the barrier separating wakefulness and anesthetic-induced unresponsiveness. *PLoS Genet.* **9**, e1003605 (2013).
34. Weber, F. & Dan, Y. Circuit-based interrogation of sleep control. *Nature* **538**, 51–59 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank all Yanagisawa/Funato laboratory and WPI-IIS members, especially T. Motoike, A. Matsui, Y. Goto, M. Takahashi and K. Taniguchi for technical assistance and their participation in the early stage of this project, M. Lazarus, R. W. Greene and K. E. Vogt for discussion and comments on this manuscript. J.S.T. is an investigator and M.Y. is a former investigator of the Howard Hughes Medical Institute. This work was supported by the World Premier International Research Center Initiative from MEXT to M.Y., JSPS KAKENHI (grant number 26220207 to M.Y., H.F., T.K.; 16K15187 to H.F.; 26507003 to C.M., H.F.; 15K18966, 00635089 to T.F.; 16K18358 to T.K.; 15J06369 to T.H.; 16K18583 to M.S.), MEXT KAKENHI (grant number; 15H05935 to M.Y., H.F.), Welch Foundation (grant number; I-1608 to Q.L.), NIH (grant number; GM111367 to Q.L.), Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST program) from JSPS to M.Y., research grants from Uehara Memorial Foundation to M.Y., and from Takeda Science Foundation to M.Y. Nematode strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We thank A. Hart and H. Huang for technical advices on nematode quiescence measurement, Y. Iino for providing plasmids, M. Ikawa for providing EGxxFP plasmid, and M. Montminy and J. B. Thomas for fly stocks.

Author Contributions H.F. and M.Y. were responsible for the overall experimental design, based on strategies conceived by M.Y. and J.S.T. M.S. constructed EEG analysis and database systems. C.M., S.K., N.H.-H., A.I., H.K., F.A., T.H., S.J.K. and K.H. conducted EEG recording and analysis. M.K. performed *in situ* hybridization. T.F., Se.M., F.S. and S.T. produced CRISPR-based gene-modified mice. M.A. and K.S. produced gene-modified mice. Sh.M., L.C. and Y.H. conducted roundworm experiments. T.K., H.M. and T.Y. conducted electrophysiological experiments. Z.W., J.M., A.W. and Q.L. conducted proteomics experiments. S.N., J.T. and K.K. conducted fruitfly experiments. ENU mice production and linkage analysis were conducted by I.M., T.S. and S.W. V.K. and J.S.T. designed B6 substrain-based screening. H.F. and M.Y. wrote the paper, which was reviewed by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Y. (yanagisawa.masa.fu@u.tsukuba.ac.jp), H.F. (funato.hiromasa.km@u.tsukuba.ac.jp) or J.S.T. (joseph.takahashi@utsouthwestern.edu)

Reviewer Information *Nature* thanks D.-J. Dijk, A. Sehgal and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animals and mutagenesis. Male C57BL/6J mice (CLEA Japan) were treated with ethylnitrosourea (85 mg kg^{-1} , Sigma-Aldrich) by intraperitoneal injection twice at weekly intervals at the age of 8 weeks. At the age of 25–30 weeks, the sperm of the mice was used for *in vitro* fertilization with eggs of C57BL/6N mice to obtain F₁ offspring. Mice were provided food and water *ad libitum*, and were maintained on a 12-h light:12-h dark cycle and housed under controlled temperature and humidity conditions. All procedures were approved by the Institutional Animal Care and Use Committee of the University of Tsukuba and the RIKEN BioResource Center, University of Texas Southwestern Medical Center at Dallas.

Surgery. EEG/EMG electrode implantation was performed as described previously³⁵, with isoflurane (3% for induction, 1% for maintenance) used for anaesthesia. Seven days after surgery, the mice were tethered to a counterbalanced arm (Instech Laboratories) that allowed free movement and exerted minimal weight.

Screening scheme. At the age of 12 weeks, male mice were implanted with EEG/EMG electrodes and then screened for sleep/wakefulness behaviour. Examined parameters were total time spent in wake, NREMS and REMS states, episode duration of wake, NREMS and REMS states, appearance of muscle atonia during REMS, and rebound sleep after 4-h sleep deprivation by shaking the cages. For quantitative parameters, we selected mice whose phenotypes deviated from the average by at least 3 standard deviations. After confirming the reproducibility of the sleep phenotype, the mice were selected for offspring production by natural mating or IVF with wild-type females to examine the heritability of the sleep phenotypes. If at least 30% of the male littermates showed sleep phenotypes similar to their father, we considered the sleep abnormalities to be heritable.

Linkage analysis. Single nucleotide polymorphisms of N₂ mice were determined using a custom TaqMan Genotyping assay (Thermo Fisher). The custom probes were designed based on the polymorphism data between C57BL/6J and C57BL/6N (ref. 20). QTL analysis was performed using J/qtl software (Jackson Laboratory).

Whole-exome sequencing. Whole exomes were captured with SureSelectXT2 Mouse All Exon (Agilent) and processed to a paired end 2×100 -bp run on the Illumina HiSeq2000 platform at the UTSW McDermott Center Next Generation Sequencing Core. Reads were mapped to the University of California Santa Cruz mm9 genome reference sequence for C57BL/6J using Burrows–Wheeler aligner and quality filtered using SAMtools. Cleaned BAM files were used to realign data and call variants using the Genome Analysis ToolKit to detect heterozygous mutations.

Sleep behaviour analysis. The recording room was kept under 12-h light:12-h dark cycles and a constant temperature (24–25°C). To examine sleep–wake behaviour under baseline conditions, EEG/EMG signals were recorded for two consecutive days from the onset of the light phase. EEG/EMG data were visualized and analysed using a MatLab (MathWorks)-based, custom semi-automated staging program followed by visual inspection. EEG signals were subjected to fast Fourier transform analysis from 1 to 30 Hz with 1-Hz bin using MatLab-based custom software. Epochs containing movement artefacts were included in the state totals but excluded from subsequent spectral analysis. Sleep/wakefulness was staged into wakefulness, NREMS and REMS. Wakefulness was scored based on the presence of low amplitude, fast EEG, and high amplitude, variable EMG. NREMS was characterized by high amplitude, delta (1–4 Hz) frequency EEG and low EMG tonus, whereas REMS was staged based on theta (6–9 Hz)-dominant EEG and EMG atonia. Hourly delta density during NREMS indicates hourly averages of delta density which is the ratio of delta power to total EEG power at each 20-s epoch. For the power spectrum of sleep/wakefulness, the EEG power of each frequency bins was expressed as a percentage of the total EEG power over all frequency bins (1–30 Hz) and sleep/wakefulness states^{35,36}. For sleep deprivation, mice were sleep deprived for 2, 4 and 6 h from the onset of the light phase by gently touching the cages when they started to recline and lower their heads. Food and water were available. To evaluate the effect of sleep deprivation, the NREMS delta power during the first hour after sleep deprivation was expressed relative to the same zeitgeber time (ZT) of the basal recording or relative to the mean of the basal recording. For caffeine and modafinil injection experiments, mice were fully acclimatized for intraperitoneal injection before sleep recording. After 24-h baseline recording, mice received caffeine (Sigma), modafinil (Sigma) or vehicle (0.5% methyl cellulose (Wako)) intraperitoneally at ZT0, followed by 12-h recording. Injections were delivered once per week, with each injection followed by a 6–8-day washout period, during which mice remained in the recording chamber. To examine the sleep/wakefulness behaviour under constant darkness, after 48-h recording under a 12-h light:12-h dark cycle, EEG/EMG recording continued in constant darkness for 3 days.

Circadian behaviour analysis. Mice were housed individually in a cage (width 23 cm, length 33 cm, height 14 cm) containing a wireless running wheel (Med Associate ENV-044). Cages were placed in a light-tight chamber equipped with

green LED light (100 lx at the bottom of the cage). The rotation numbers of wheels were obtained with 1-min bin using Wheel manager software (Med Associate). Mice were entrained to 12-h light:12-h dark cycle for 7 days, and then released into constant darkness for 3 weeks. The free running period was calculated with linear regression analysis of activity onset using MatLab-based custom software. Circadian activity amplitude was calculated by fast Fourier transform of activity data, which were processed with Bartlett window using MatLab-based custom software. Relative amplitude was normalized to the mean amplitude of the wild-type group.

Western blot. A rabbit polyclonal antibody against the C-terminal 171 amino acids of mouse SIK3 was generated using custom antibody production service (Pacific Immunology). Tissues were homogenized using a rotor-stator homogenizer (Polytron) in ice-cold lysis buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM Na₂P₂O₇, 1.5% Triton X-100, 15 mM NaF, $1 \times$ PhosSTOP (Roche), 5 mM EDTA, $1 \times$ protease inhibitor (Roche)), and then centrifuged at $13,000g$ at 4°C. The supernatants were separated by SDS–PAGE and transferred on PVDF membrane. Western blotting was performed according to standard protocols.

In situ hybridization. *In situ* hybridization was performed as described previously³⁷. In brief, a 0.7–0.8-kb fragment of *Nalcn* cDNA was inserted into pGEM-T easy (Promega) and used for DIG-labelled probe synthesis. Mice were deeply anaesthetized with sodium pentobarbital and perfused transcardially with PBS followed by 4% paraformaldehyde (PFA). Forty-micrometre-thick brain sections were treated with 0.3% Triton X-100, digested with $1 \mu\text{g ml}^{-1}$ proteinase K, treated with 0.75% glycine, and then treated with 0.25% acetic anhydride in 0.1 M triethanolamine. After overnight incubation with a digoxigenin (DIG)-labelled probe at 60°C, the sections were washed and then incubated with alkaline phosphatase-conjugated anti-DIG Fab fragments (Roche, 11175041910). The reactions were visualized with a 5-bromo-4-chloro-3-indolyl-phosphate/4-nitroblue tetrazolium (BCIP/NBT) substrate solution (Roche).

Cell lines. HEK293 cells (RCB1637) and HEK293T cells (RCB2202) were obtained from the RIKEN BRC Cell Bank. Cells were cultured in DMEM (Wako) supplemented with 10% FBS, 1% GlutaMAX (Thermo Fisher Scientific), and penicillin/streptomycin at 37°C in a humidified atmosphere of 5% CO₂. Cell lines were regularly tested for mycoplasma contamination using MycoAlert (Lonza). Cell lines were regularly renewed by obtaining cell stocks from the Cell Bank for authentication. We used HEK293 and HEK293T cells because of their reliable growth, high efficiency in transfection and morphology suitable for electrophysiological experiments.

Production of *Sik3* gene-modified mice by conventional gene targeting. For generating *Sik3*^{Slp} knock-in mice, a genomic fragment containing exon 13 of the *Sik3* gene was isolated from C57BL/6 mouse genomic BAC clone from a RP23 mouse genomic BAC library (Advanced Genetechnics Co). A 1.7-kb fragment of FRT-PGK-gb2-neo-FRT-loxP cassette (Gene Bridges) flanked by two flippase recognition target (FRT) sites was inserted before exon 12. The targeting vector also contains a G-to-A substitution at the fifth nucleotide from the beginning of intron 13. The targeting vector was linearized and electroporated into the C57BL/6N ES cell line RENKA. Correctly targeted clones were injected into eight-cell stage ICR mouse embryos, which were cultured to produce blastocysts and then transferred to pseudopregnant ICR females. Resulting male chimaeric mice were crossed with female C57BL/6N mice to establish the *Sik3*^{Slp-neo/+} line. To remove the neomycin resistance gene with the FLP-FRT system, *Sik3*^{Slp-neo/+} mice were crossed with *Actb-FLP* knock-in mice.

Production of *Sik3* gene-modified mice using ZFN. The custom-designed ZFN mRNAs targeting the exon 13–intron 13 boundary region of the *Sik3* gene were obtained from Sigma-Aldrich's Composers Custom ZFN service. Before the final assembly of the ZFN products, Sigma-Aldrich validated the designed ZFN binding sequences *in silico* using their bioinformatics tools and *in vitro* using Nero2A cell lines, ensuring high cutting efficiency and specificity using mismatch-specific endonuclease Cell according to the manufacturer's instructions. The ZFN mRNAs were injected into single-cell stage C57BL/6J mouse zygotes at the University of Texas Southwestern Transgenic Core facility. The injected eggs were then transferred to pseudopregnant females to generate F₀ founders. In total, 45 out of 96 F₀ mice were found to be modified at the exon 13–intron 13 boundary region of the *Sik3* gene. We crossed one F₀ male mouse that had a 2-bp deletion from the last nucleotide of exon 13 with female C57BL/6N mice to obtain F₁ mice of *Sik3*^{Slp/+} ZFN. The F₁ mice were used to confirm the skipping of exon 13 in *Sik3* mRNA, which was purified from the brains and livers. The F₂ male mice were used for sleep/wakefulness behaviour analysis.

Production of *Flag-Sik3* mice and *Nalcn*^{DRL} mice by CRISPR/Cas9 technology. To produce a Cas9/single-guide RNA (sgRNA) expression vector, oligonucleotide DNAs (5'-CACCGCGAGCGGCCATCGACCCGC-3' and 5'-AAACGCGGGTCGATGGCCGCTCGC-3') were annealed and then inserted into pX330

vector (Addgene). The cleavage activity of the pX330-Sik3Ex1 vector was evaluated by the EGFP system³⁸. Genomic DNA containing exon 1 of the *Sik3* gene was amplified and inserted into pCAG-EGFP to produce pCAG-EGFP-Sik3Ex1. The pX330-Sik3Ex1 and pCAG-EGFP-Sik3Ex1 were transfected into HEK293 cells. As a donor oligonucleotide, a single-stranded 200-nucleotide DNA was synthesized (Integrated DNA Technologies), which contained a Flag-haemagglutinin-coding sequence in the centre and 70-nucleotide arms at the 5' and 3' ends. Female C57BL/6J mice or *Sik3^{Slp}* knock-in mice were injected with pregnant mare serum gonadotropin and human chorionic gonadotropin at a 48-h interval, and mated with male C57BL/6J mice. The fertilized one-cell embryos were collected from the oviducts. Then, 5 ng μl^{-1} of pX330-Sik3Ex1 vector and 10 ng μl^{-1} of the donor oligonucleotide were injected into the pronuclei of these one-cell-stage embryos. The injected one-cell embryos were then transferred into pseudopregnant ICR mice. F₀ mice were genotyped for the presence of Flag-coding sequence in exon1 of the *Sik3* gene and for the presence of the *Sik3^{Slp}* mutation. F₀ mice containing Flag-SIK3 were further examined for the presence of the Cas9 transgene and off-target effects. Candidate off-target sites were identified based on a complete match of 16 bp at the 3' end, including the PAM sequence. F₀ mice were mated with C57BL/6N mice to obtain F₁ offspring.

Nalcn^{Dr1} mice were produced as described above. To produce the sgRNA expression vector, pX330-NalcnEx9, oligonucleotide DNAs (5'-CACCA GCAATAAACACATTCTGAA-3' and 5'-AAACTTCAGATGTGTGTTA TTGCT-3') were used. Genomic DNA containing exon 9 of the *Nalcn* gene was amplified and inserted into pCAG-EGFP to produce pCAG-EGFP-NalcnEx9. As a donor oligonucleotide, a single-stranded 199-nucleotide DNA containing a T-to-A substitution at the centre was synthesized (Integrated DNA Technologies). *Nalcn* mutant mice of N₂-N₃ generation were used for sleep/wakefulness analysis. **Phosphoproteomic analysis.** To evaluate Flag-tagged SIK3 protein in brains, we performed peptide mapping of the purified Flag-SIK3 protein. The brains of *Flag-Sik3* knock-in mice and *Flag-Sik3^{Slp}* knock-in mice were quickly dissected after cervical dislocation. Brains were homogenized in detergent-free buffer and then centrifuged (100,000g, 30 min, 4°C). The supernatant was immunoprecipitated with anti-DDDDK antibody beads (MBL 3325). The eluate was run on a polyacrylamide gel and stained with SilverQuest Silver staining kit (Life technologies). Flag-SIK3 band (150 kDa) was dissected with a fresh blade. The proteins in the bands were reduced with 10 mM dithiothreitol and alkylated with 40 mM iodoacetamide. Each sample was digested with trypsin (4 $\mu\text{g ml}^{-1}$; Trypsin Gold, Promega) at 37°C overnight. The extracted peptides were then separated via nano flow LC (Advance LC, Michrom Bioresources) using a C18 column. The LC eluate was coupled to a nano-ion-spray source attached to a Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). All MS/MS spectra were searched using Proteome Discoverer 1.3 software (Thermo Fisher Scientific). Peptides were mapped through mouse SIK3 (NP_081774) with 56% coverage.

To examine the effect of sleep deprivation on the phosphorylation status of SIK3 protein, five *Flag-Sik3* knock-in mice or five *Flag-Sik3^{Slp}* knock-in mice were ad libitum slept (S) or sleep-deprived (SD) for 4 h by gentle handling immediately after light onset (ZT0–ZT4). Five wild-type mice were used as a negative control. At ZT4, mouse brains were quickly dissected after cervical dislocation, rinsed with cold PBS, and snap frozen in liquid nitrogen. Each half of the brains was lysed in 2 ml of ice-cold lysis buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 2 mM MgCl₂, 15 mM NaF, 10 mM Na₄P₂O₇) freshly supplemented with protease/phosphatase inhibitor cocktail tablets (Roche), and homogenized in a glass tissue homogenizer. After brain homogenate was incubated for 30 min and centrifuged at 13,000g for 20 min at 4°C, the supernatant was pre-cleared by IgG and Protein G beads for 30 min before immunoprecipitation. Each pre-cleared lysate was added to 50 μl of anti-Flag antibody-conjugated Sepharose beads (Sigma, A2220) and rotated overnight at 4°C. After washing the beads five times with cold wash buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 2 mM MgCl₂, 15 mM NaF, 10 mM Na₄P₂O₇), 50 μl of elution buffer (2% SDS, 60 mM Tris-HCl, pH 6.8, 50 mM DTT, 10% glycerol) was added and rotated for 10 min at 4°C. Elution was repeated twice and combined into one eluate and analysed by western blotting. For each group of *Flag* knock-in mice, the five eluates of were mixed and equally split into two or three samples for mass spectrometric analysis. Thus, a total of six (*Flag-Sik3*) or nine (*Flag-Sik3* and *Flag-Sik3^{Slp}*) samples were reduced, alkylated, and trypsin digested overnight. After desalting, each sample was labelled with a different Tandem Mass Tag (TMT) reagent (Thermo Fisher Scientific), then all samples were combined into one mixture for HPLC fractionation using a C18 column. A total of 12 fractions were collected, and analysed separately on the Orbitrap-Fusion mass spectrometry platform (Thermo Fisher Scientific) using a reverse-phase liquid chromatography tandem mass spectrometry (LC-MS/MS) method. We performed data analysis to identify peptides and quantified reporter ion relative abundance using Proteome

Discoverer 2.1 (Thermo Fisher Scientific). The relative abundance of quantified SIK3 phosphorylation sites was normalized with wild-type negative control and total SIK3 protein abundance.

Patch-clamp recordings from HEK cells. To express wild-type NALCN, we used pTracer-CMV2-ratNALCN-EF1 α -EGFP (a gift from D. Ren)³⁹. A single nucleotide substitution was induced to make pTracer-CMV2-ratNALCN(DRL)-EF1 α -EGFP using a KOD plus Mutagenesis kit (Toyobo). HEK293T cells were grown to ~50% confluency in 12-well plates. Using Lipofectamine LTX (2 μl) and PLUS (1 μl) reagents (Thermo Fisher Scientific), the cells were cotransfected with 0.3 μg of each plasmid DNA encoding rat NALCN-EGFP (wild type or DRL), mouse UNC-80, and mouse SRC (Y529F) (constitutively active Src) in 12-well plates. UNC-80 and SRC kinase activate NALCN^{27,28}. In some experiments, the cells were incubated with 10 μM Gd³⁺ to inhibit NALCN. The cells were dissociated and plated on 18-mm coverslips coated with poly-L-lysine in fresh culture medium before patch-clamp recordings.

All patch-clamp recordings from HEK293T cells were performed >72 h after transfection. Recording patch pipettes were pulled from glass capillaries (1B150F-4, World Precision Instruments) using a micropipette puller (P-97, Sutter Instrument) to give a resistance of ~9 M Ω . The series resistance of whole-cell recordings was ~40 M Ω , which was not compensated. Patch pipettes were filled with solution containing 150 mM CsOH, 120 mM methanesulfonic acid, 10 mM NaCl, 10 mM EGTA, 2 mM Mg₂ATP and 10 mM HEPES (pH 7.4 adjusted with methanesulfonic acid; osmolality, 290–299 mOsm l⁻¹ adjusted with CsCl). The cells on coverslips were transferred to a recording chamber under a fluorescence upright microscope (Axio Examiner D1, Zeiss) and continuously perfused with the bath solutions containing 150 mM NaCl, 3.5 mM KCl, 10 mM HEPES, 20 mM glucose, 5 mM NaOH, 2 mM MgCl₂ and 1.2 mM CaCl₂ (pH 7.4 adjusted; osmolality, 300–310 mOsm l⁻¹). The transfected cells were identified by enhance green fluorescent protein (eGFP) fluorescence. Patch-clamp recordings were performed at room temperature (24°C) using a computer-controlled amplifier (MultiClamp 700B, Molecular Devices). The signals were digitized with A/D converter (Digidata 1440A, Molecular Devices), and acquired with Clampex (Molecular Devices) at a sampling rate of 50 kHz, and low-pass filtered at 5 kHz. At the end of recording, Gd³⁺ (10 μM) was used to confirm that the whole-cell currents were mediated through NALCN³⁹. Data were analysed using Clampfit (Molecular Devices). The equilibrium potentials were calculated from *I*-*V* curves. Mean membrane conductance was estimated from the regression lines fitted to *I*-*V* curves from individual cells. Current, membrane conductance and charge transfer were normalized to membrane capacitance.

Patch-clamp recordings from neurons. Patch pipettes and recording system were the same as those used in recordings from HEK293 cells. Acute brain slices containing the DpMe were prepared from post-natal day 12–23 *Nalcn^{+/+}* or *Nalcn^{Dr1/+}* mice. After the induction of deep anaesthesia with isoflurane, mice were decapitated and the brains were rapidly removed into an ice-cold cutting solution containing 2.5 mM KCl, 1.25 mM NaH₂PO₄, 26 mM NaHCO₃, 25 mM glucose, 185 mM sucrose, 0.5 mM CaCl₂ and 10 mM MgCl₂ (pH 7.4, when bubbled with 95% O₂ and 5% CO₂). The brains were cut coronally into 200–250 μm -thick slices with a vibratome (VT-1200S, Leica). The slices were incubated at 37°C for 1 h in artificial cerebrospinal fluid (aCSF) containing 125 mM NaCl, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 26 mM NaHCO₃, 10 mM glucose, 2 mM CaCl₂ and 1 mM MgCl₂ (pH 7.4, when bubbled with 95% O₂ and 5% CO₂) before recordings. Slices were transferred to a recording chamber perfused with aCSF under an upright microscope (Axio Examiner D1, Zeiss). Patch pipettes were filled with solution containing 125 mM K-gluconate, 10 mM KCl, 10 mM HEPES, 0.5 mM EGTA, 8 mM phosphocreatine-Na₂, 4 mM ATP-Mg and 0.3 mM GTP-Na (pH 7.3 adjusted with KOH; osmolality, 290 mOsm l⁻¹). The DpMe was identified with axon bundles. Recordings were made from cells located in the medial part of the DpMe. Cells showing no action potentials following current injection (>1 nA, >5 ms) were discarded from analysis. Membrane potentials were recorded for 1–10 min.

Fruitfly stocks and behavioural assay. *Sik3* hypomorph and *UAS-Sik3*, *UAS-Sik3(S563A)* transgenic flies were gifts from M. Montminy and J. B. Thomas⁴⁰. *elav-GS* (GeneSwitch) stocks were from the Bloomington stock centre. Flies were reared at 25°C under 12-h light:12-h dark cycle in 50–60% relative humidity on a standard fly food consisting of corn meal, yeast, glucose, wheat germ and agar.

Sleep analysis was performed as described previously⁴¹. In brief, male flies (2–5 days old) were individually housed in glass tubes (length, 65 mm; inside diameter, 3 mm) containing standard fly food at one end and a cotton plug on the other end. Sucrose-agar (1% agar supplemented with 5% sucrose) food was used for the GeneSwitch system assay, instead of standard food. The glass tubes were placed in the *Drosophila* activity monitor (DAM) (Trikinetics) and the locomotor activity of each fly was recorded as the number of infrared beam crossings in 1-min bin. Sleep was defined as periods of inactivity lasting 5 min or longer. Sleep assay were performed for 3 d under 12-h light:12-h dark cycle conditions and then constant

darkness conditions. For 12-h light:12-h dark cycles, zeitgeber time (ZT) was used, and for constant darkness, circadian time (CT), with CT0 as 12 h after lights-off of the last 12-h light:12-h dark conditions, was used to indicate the daily time.

For conditional expression analysis, we used the GeneSwitch system⁴² where expression is induced by a steroid hormone antagonist RU486. Flies are monitored for 3 days in tubes without drug in constant darkness and then transferred to new tubes either with vehicle (0.5% DMSO) alone or with 0.5 mM RU486 and then further monitored under constant darkness conditions. The expression of endogenous or transgenic *Sik3* genes was confirmed by RT-PCR using RNA from fly heads.

Nematode strains and quiescence assay. The wild-type strain N₂ and the mutant strain PY1479 *kin-29(oy38)* X were obtained from the *Caenorhabditis* Genetics Center (CGC)⁴³. All worms were maintained at 20 °C on nematode growth medium (NGM) agar plates seeded with *E. coli* HB101. For construction of *P_{H20::kin-29}*, *kin-29* cDNA was amplified by RT-PCR and inserted into the plasmid pPD-DEST (a gift from Y. Iino) to generate pDEST-KIN-29. Next, we carried out the LR-recombinase reaction (Gateway System, Life Technologies) between pENTR-*P_{H20}* (a gift from Y. Iino) and pDEST-KIN-29 to generate *P_{H20::kin-29}*. *P_{H20::kin-29}* was injected at 30 ng μl^{-1} together with the injection marker *P_{myo-3::mcherry}* (10 ng μl^{-1}) and the empty vector pPD49_26 (60 ng μl^{-1}) into the *kin-29(oy38)* mutant worms.

Quiescence during the L4 to adult lethargus was measured using the microfluidic-chamber based assay⁴⁴. In brief, polydimethylsiloxane-made microfluidic chambers containing liquid NGM and the *E. coli* HB101 were loaded with early L4 larvae and sealed with a cover glass plus 2% agarose, and set under the microscope. Images were taken every 2 s for 12 to 20 h at 20 \pm 0.5 °C using the microscope M205FA (Leica) equipped with the camera MC120HD (Leica) (pixel size: 1,024 $\mu\text{m} \times$ 768 μm) controlled by Leica Application Suite V4.3 or the microscope SZX16 (Olympus) equipped with the camera GR500BCM2 (Shodensha) (pixel size: 1,024 $\mu\text{m} \times$ 768 μm) controlled by μ Manager (UCSF). Subtraction between serial images was carried out using Image J, and worms were regarded as quiescent at a specific time point if the difference from the preceding time point was less than 1% of the total body size. The fraction of quiescence was defined as the number of quiescent time points divided by the total number of time points during a period of 10 min.

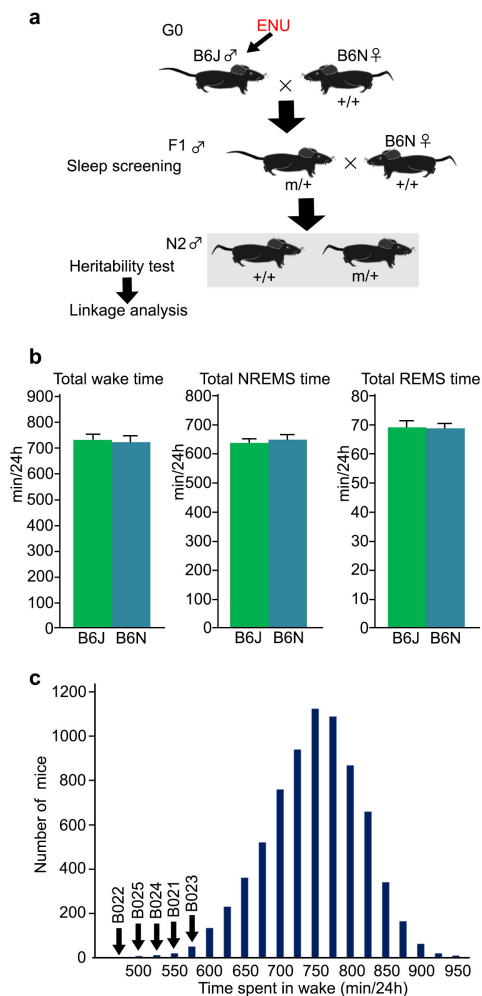
The onset of lethargus quiescence was defined as the time point after which the fraction of quiescence was higher than 0.05 for at least 20 min, whereas the end point was defined as the time point after which the fraction of quiescence was lower than 0.05 for at least 20 min. Occasionally, brief episodes of quiescence were observed outside of lethargus both in wild-type and mutant worms; these

episodes were excluded by setting a threshold of 60 min for the minimum duration of lethargus quiescence.

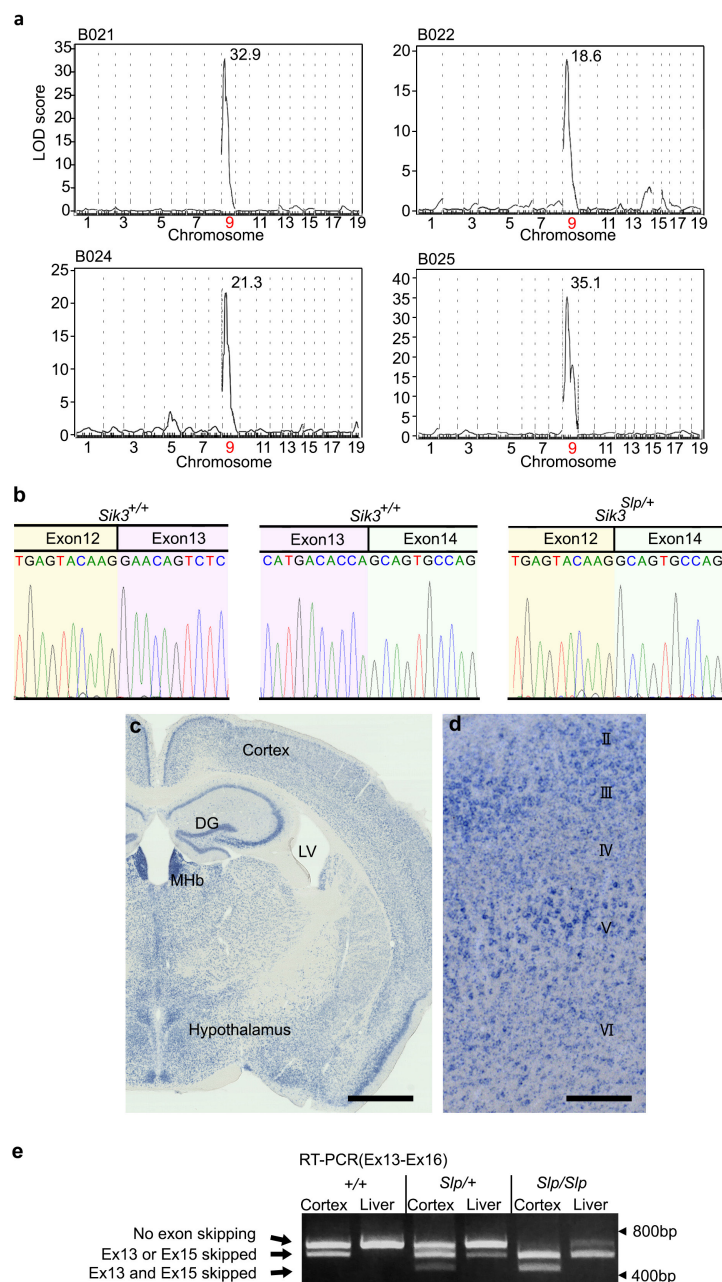
Statistics. Sample sizes were determined using R software based on averages and standard deviations that were obtained from small scale experiments. No method of randomization was used in any of the experiments. The experimenters were blinded to genotypes and treatment assignment. Statistical analysis was performed using SPSS Statistics 22 (IBM) and R software. All data were tested for Gaussian distribution and variance. Homogeneity of variances was tested with Levene's test. We used Student's *t*-test for pairwise comparisons, one-way ANOVA for multiple comparisons, one-way repeated measure ANOVA for multiple comparisons with multiple data points, and two-way ANOVA for multiple comparisons involving two independent variables. ANOVA analyses were subjected to Tukey's post-hoc test. When deviation from normality and lack of homogeneity of variances occurred ($P < 0.05$), Mann-Whitney *U* test was used for group comparison. $P < 0.05$ was considered statistically significant.

Data availability. The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

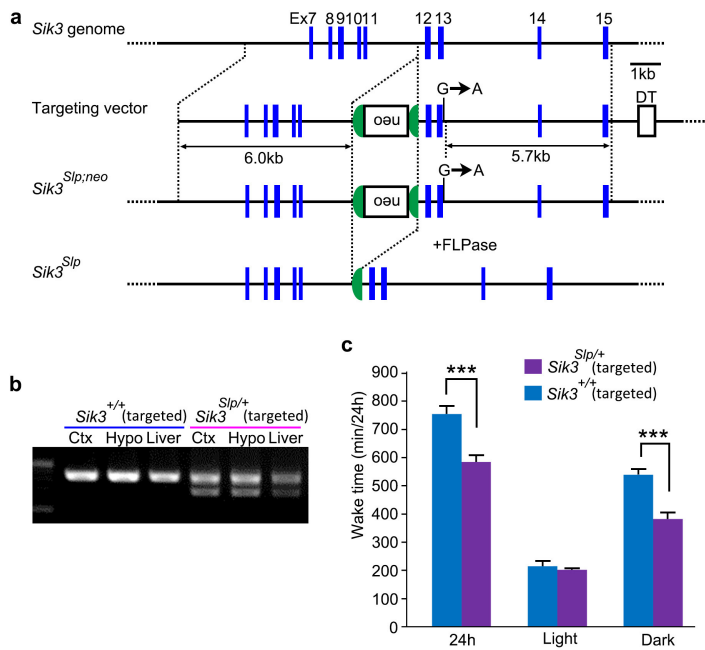
35. Funato, H. *et al.* Loss of Goosecoid-like and DiGeorge syndrome critical region 14 in interpeduncular nucleus results in altered regulation of rapid eye movement sleep. *Proc. Natl Acad. Sci. USA* **107**, 18155–18160 (2010).
36. Franken, P., Malafosse, A. & Tafti, M. Genetic variation in EEG activity during sleep in inbred mice. *Am. J. Physiol.* **275**, R1127–R1137 (1998).
37. Funato, H., Saito-Nakazato, Y. & Takahashi, H. Axonal growth from the habenular nucleus along the neuromere boundary region of the diencephalon is regulated by semaphorin 3F and netrin-1. *Mol. Cell. Neurosci.* **16**, 206–220 (2000).
38. Mashiko, D. *et al.* Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci. Rep.* **3**, 3355 (2013).
39. Lu, B. *et al.* The neuronal channel NALCN contributes resting sodium permeability and is required for normal respiratory rhythm. *Cell* **129**, 371–383 (2007).
40. Wang, B. *et al.* A hormone-dependent module regulating energy balance. *Cell* **145**, 596–606 (2011).
41. Kume, K., Kume, S., Park, S. K., Hirsh, J. & Jackson, F. R. Dopamine is a regulator of arousal in the fruit fly. *J. Neurosci.* **25**, 7377–7384 (2005).
42. Osterwalder, T., Yoon, K. S., White, B. H. & Keshishian, H. A conditional tissue-specific transgene expression system using inducible GAL4. *Proc. Natl Acad. Sci. USA* **98**, 12596–12601 (2001).
43. Lanjuin, A. & Sengupta, P. Regulation of chemosensory receptor expression and sensory signaling by the KIN-29 Ser/Thr kinase. *Neuron* **33**, 369–381 (2002).
44. Singh, K. *et al.* *C. elegans* Notch signaling regulates adult chemosensory response and larval molting quiescence. *Curr. Biol.* **21**, 825–834 (2011).



Extended Data Figure 1 | Sleep/wakefulness screening of randomly mutagenized mice. **a**, ENU-treated G_0 mice were mated with B6N females to obtain the offspring. The F_1 mice were used for sleep/wakefulness analysis. A mouse showing any sleep abnormalities was crossed with B6N female mice. The N_2 progeny was examined for heritability of sleep abnormality and for chromosomal mapping. **b**, B6J ($n = 20$) and B6N ($n = 21$) showed similar total wake time (left, $P = 0.67$, two-tailed Student's t -test), NREMS time (centre, $P = 0.66$) and REMS time (right, $P = 0.84$). Data are mean \pm s.e.m. **c**, The histogram shows total daily wake time of all mice screened. Total wake time of screened mice was 735 ± 66.9 min (mean \pm s.d.). Arrows indicate the founders of *Sleepy* mutant pedigrees.

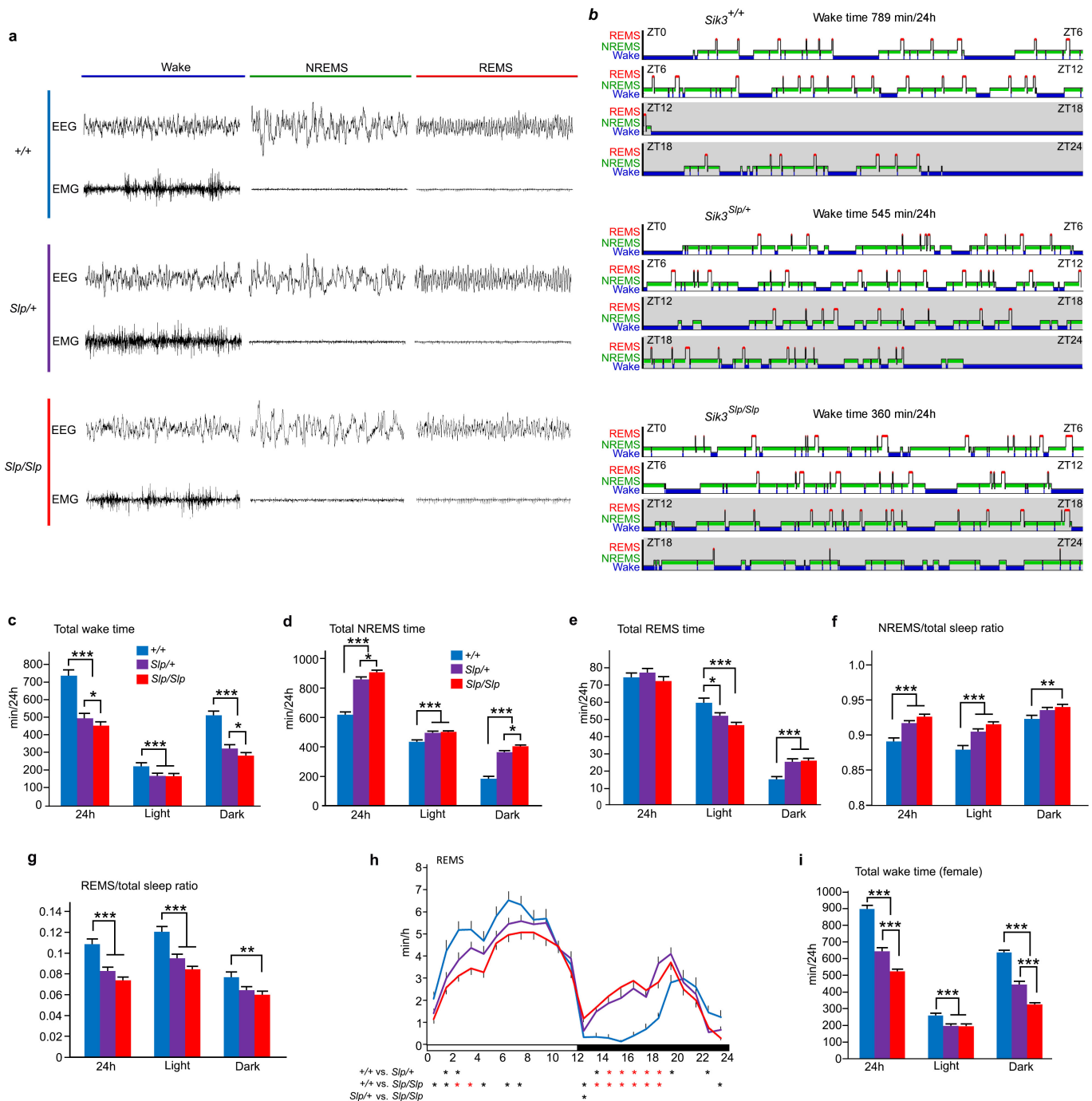


Extended Data Figure 2 | QTL analysis of *Sleepy* mutant pedigrees and characterization of *Sik3* transcript. **a**, QTL analysis of B021 ($n = 119$), B022 ($n = 95$), B024 ($n = 59$) and B025 ($n = 112$) pedigrees for total wake time produced a single LOD score peak on chromosome 9. **b**, Direct sequencing of the exon 12/13 boundary and exon 13/14 boundary of *Sik3* mRNA of *Sik3*^{+/+} mouse. Direct sequencing of the short RT-PCR product specific to *Sik3* mutant mice shows the direct transition from exon 12 to exon 14. **c**, **d**, *Sik3* mRNA is expressed broadly in forebrain neurons (**c**). *Sik3* mRNA is expressed throughout the cerebral cortex in the primary motor area (**d**). DG, dentate gyrus; LV, lateral ventricle; MHb, medial habenula. Scale bars, 1 mm (**c**) and 250 μ m (**d**). **e**, RT-PCR of *Sik3* mRNA from cerebral cortex and liver of *Sik3*^{+/+}, *Sik3*^{Slp/+} and *Sik3*^{Slp/Slp} mice. Normal *Sik3* variant lacking exon 15 expressed in the cerebral cortex.



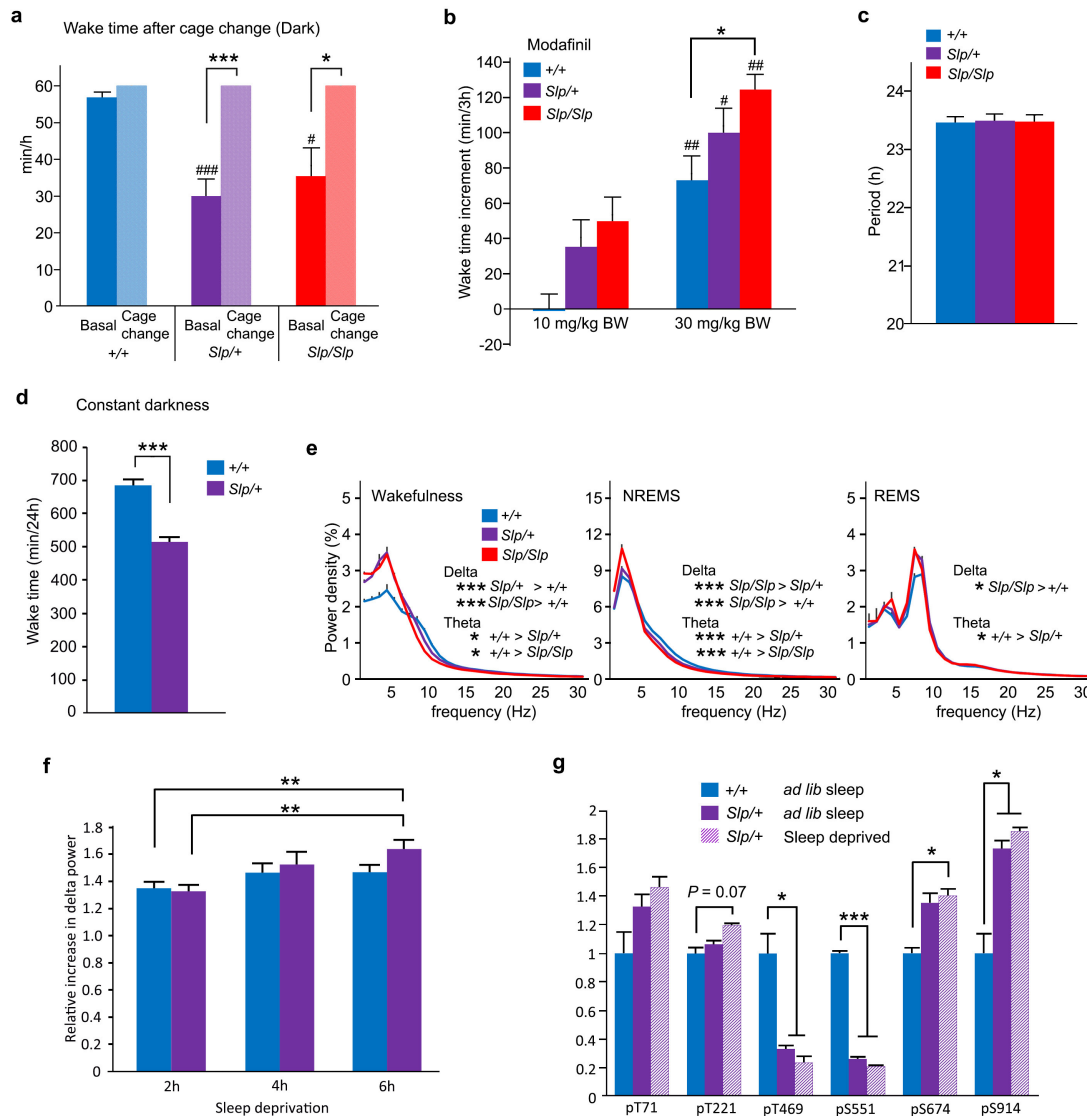
Extended Data Figure 3 | Sleep/wakefulness of *Sik3*^{Slp} knock-in mice.

a, The structure of the *Sik3* genome and targeting vector for *Sik3*^{Slp}. Neomycin resistance gene under the mouse phosphoglycerol kinase promoter (*neo*) was sandwiched with the flippase recognition target (FRT) sequences. The guanine at the fifth nucleotide from the beginning of the intron 13 was substituted with adenine. The *neo* cassette was deleted by crossing with *Actb*^{CAG-FLP} knock-in mice. **b**, RT-PCR of *Sik3* mRNA of *Sik3*^{Slp/+} knock-in mice. **c**, Total wake time of *Sik3*^{Slp/+} knock-in mice ($n = 10$) and *Sik3*^{+/+} littermates ($n = 6$). *** $P < 0.001$, two-way ANOVA followed by Tukey's test. Data are mean \pm s.e.m.



Extended Data Figure 4 | Sleep/wakefulness behaviours of *Sik3* mutant mice. **a**, Representative 8-s EEG and EMG for wake, NREMS and REMS of *Sik3* mutant mice. **b**, Representative hypnogram of *Sik3* mutant mice. Wake (blue), NREMS (green) and REMS (red) are indicated from ZT0 to ZT24. **c–g**, Total wake time (**c**), NREMS time (**d**), REMS time (**e**), NREMS/total sleep ratio (**f**), REMS/total sleep ratio (**g**) and circadian variation of REMS (**h**) of *Sik3*^{+/+} (*n* = 22), *Sik3*^{Slp/+} (*n* = 32) and *Sik3*^{Slp/Slp}

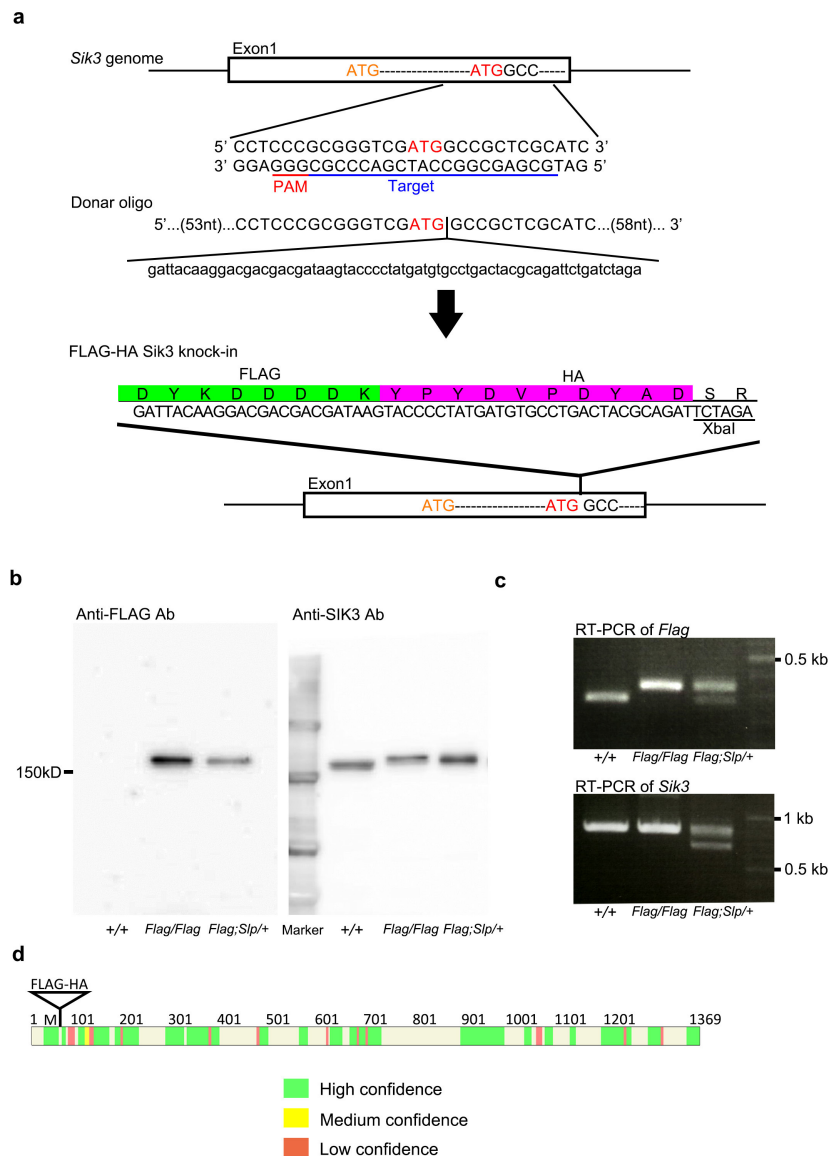
(*n* = 31) mice. **P* < 0.05; ***P* < 0.01; ****P* < 0.001, two-way ANOVA followed by Tukey's test (**c–g**). **P* < 0.05 (red); **P* < 0.001 (black), one-way repeated measures ANOVA followed by Tukey's test (**h**). **i**, Total wake time of female *Sik3*^{+/+} (*n* = 10), *Sik3*^{Slp/+} (*n* = 11) and *Sik3*^{Slp/Slp} (*n* = 9) mice. ****P* < 0.001, two-way ANOVA followed by Tukey's test. Data are mean ± s.e.m.



Extended Data Figure 5 | Characterization of sleep/wakefulness

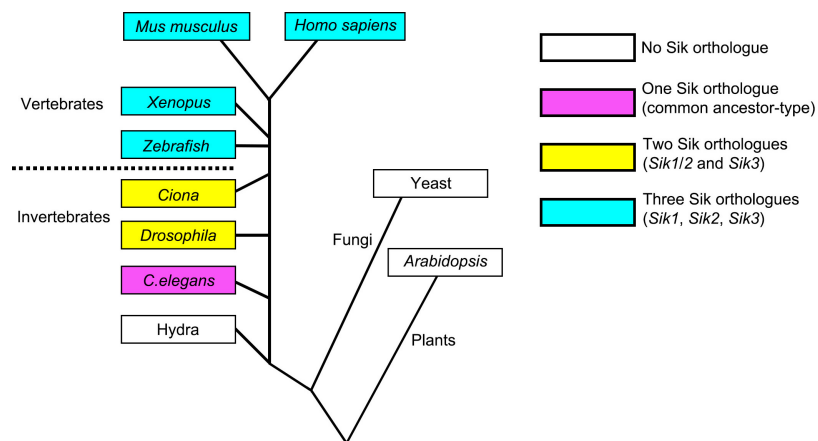
behaviours of *Sik3* mutant mice. **a**, Wake time after cage change at ZT15 in *Sik3*^{+/+} ($n = 5$), *Sik3*^{Slp/+} ($n = 10$) and *Sik3*^{Slp/Slp} ($n = 5$) mice. The graph shows time spent in wakefulness from ZT15 to ZT16 under a basal condition and after cage change from the home cage to a new cage at ZT15. * $P < 0.05$; *** $P < 0.001$ versus *Sik3*^{+/+}; # $P < 0.05$; ### $P < 0.001$, one-way repeated measures ANOVA followed by Tukey's test. **b**, Wake time increases for 3 h after modafinil injection at ZT0 in *Sik3*^{+/+} ($n = 6$), *Sik3*^{Slp/+} ($n = 6$) and *Sik3*^{Slp/Slp} ($n = 6$) mice. * $P < 0.05$; versus modafinil 10 mg kg⁻¹ in the same genotype, # $P < 0.05$, ## $P < 0.01$, two-way ANOVA followed by Tukey's test. **c**, The circadian period under constant darkness in *Sik3*^{+/+} ($n = 8$), *Sik3*^{Slp/+} ($n = 8$) and *Sik3*^{Slp/Slp} ($n = 6$) mice. $P = 0.97$,

one-way ANOVA. **d**, Total wake time of *Sik3*^{+/+} ($n = 9$) and *Sik3*^{Slp/+} ($n = 12$) mice under constant darkness. *** $P < 0.001$, two-tailed Student's t -test. **e**, EEG power spectra of *Sik3*^{+/+} ($n = 22$), *Sik3*^{Slp/+} ($n = 32$) and *Sik3*^{Slp/Slp} ($n = 31$) mice. * $P < 0.05$; *** $P < 0.001$, one-way ANOVA followed by Tukey's test. **f**, Increase in NREMS delta power after 2 h, 4 h and 6 h of sleep deprivation of *Sik3*^{+/+} ($n = 11$) and *Sik3*^{Slp/+} ($n = 11$) mice relative to mean NREMS delta power during basal sleep. ** $P < 0.01$, two-way ANOVA followed by Tukey's test. **g**, Phosphorylation of Flag-SIK3 of *Flag-Sik3*^{+/+} brains and of *Flag-SIK3*(SLP) of *Flag-Sik3*^{Slp/+} brains with or without 4-h sleep deprivation. * $P < 0.05$; *** $P < 0.001$, two-way ANOVA followed by Tukey's test. Data are mean \pm s.e.m.

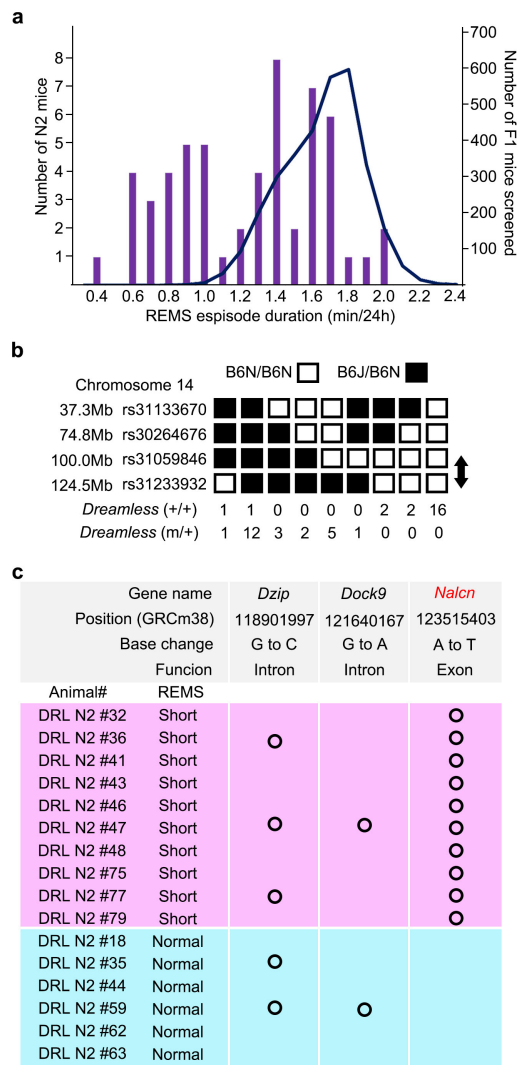


Extended Data Figure 6 | Characterization of *Flag-Sik3* mice made by CRISPR/Cas9 technology. **a**, Exon 1 of the *Sik3* gene contains the first and second methionine residues. The single-guide RNA was designed to target the second methionine-coding region. The donor oligonucleotide has a Flag-haemagglutinin (HA)-coding sequence immediately after the second methionine and 70-nucleotide long arms at both 5' and 3' ends. The Flag-HA-coding region is followed by an XbaI site. **b**, Immunoblotting of brain homogenates of *Sik3*^{+/+}, *Sik3*^{Flag/Flag} *Sik3*^{Flag.Slp/+} mice showed

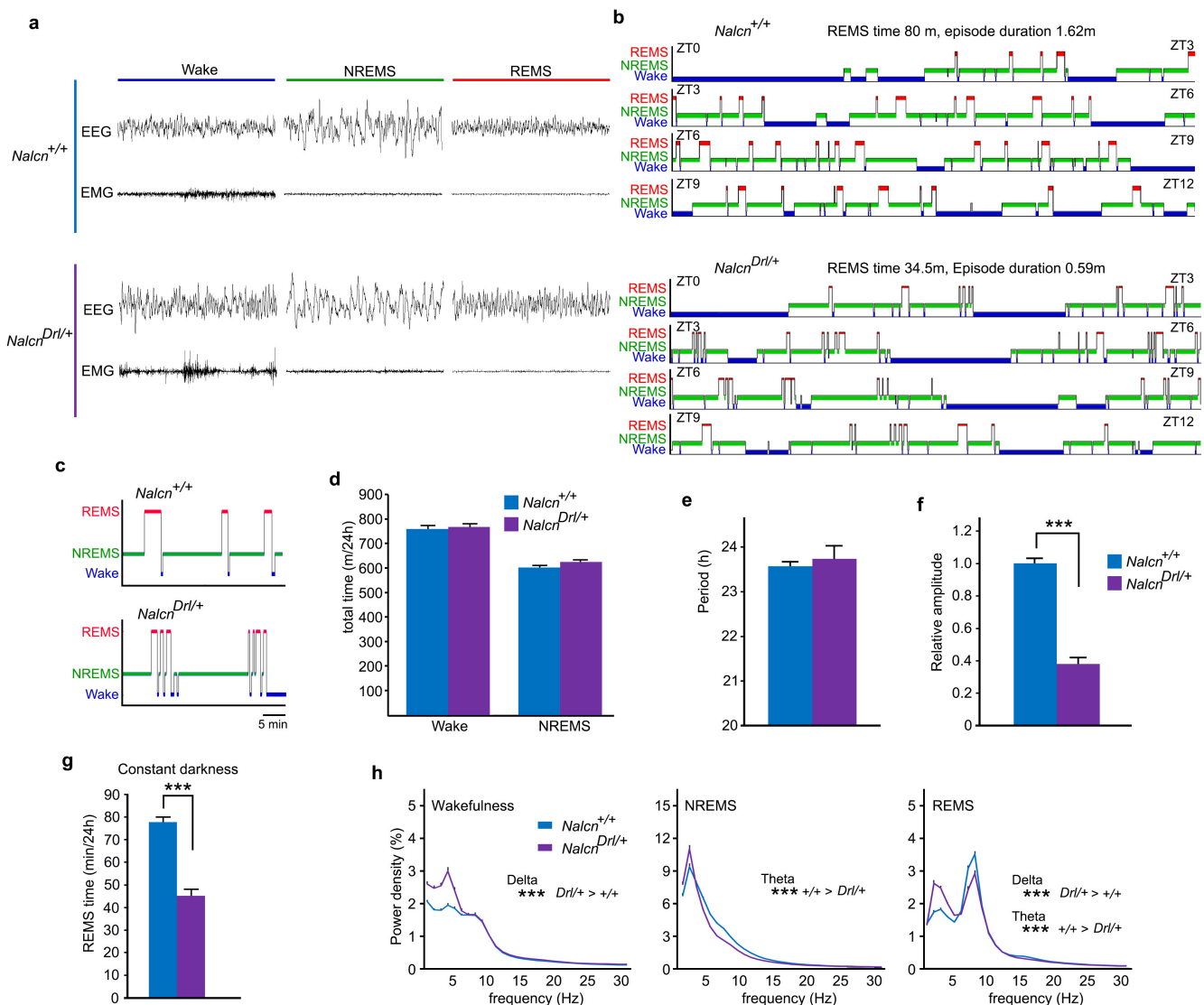
that anti-Flag antibody detected Flag-SIK3 protein of *Sik3*^{Flag/Flag} brains and Flag-SIK3 (SLP) protein of *Sik3*^{Flag.Slp/+} brains, whereas anti-SIK3 antibody detected SIK3 proteins of all genotypes. **c**, RT-PCR of brain *Sik3* mRNA of *Sik3*^{+/+}, *Sik3*^{Flag/Flag}, *Sik3*^{Flag.Slp/+} mice. **d**, Tryptic peptides of immunoprecipitated and gel-purified Flag-SIK3 protein were analysed by LC-MS and mapped on the reference SIK3 protein. The peptide fragments were mapped on almost entire SIK3 protein with high confidence.



Extended Data Figure 7 | Phylogenetic conservation of the SIK3 protein.

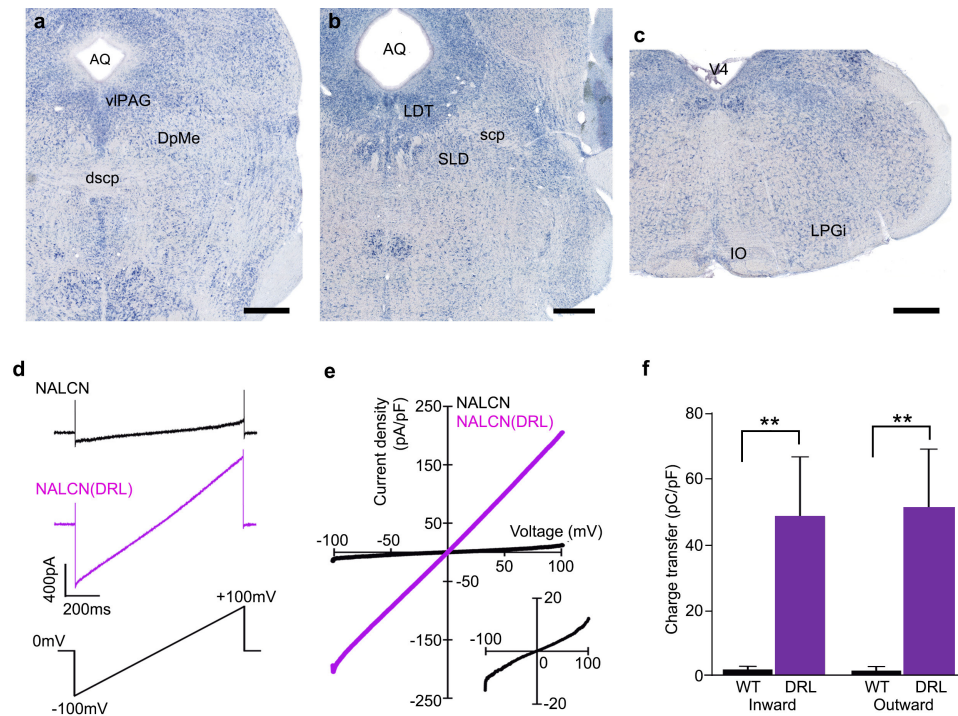


Extended Data Figure 8 | Identification of *Nalcn* mutation of the *Dreamless* mutant pedigree. **a**, Histogram of REMS episode duration in N₂ littermates of *Dreamless* mutant pedigree (bars) and all F₁ mice examined (curve). **b**, Haplotype analysis of chromosome 14 of *Dreamless* mutant pedigree with or without short REMS episode duration. **c**, Whole-exome sequencing of *Dreamless* mutant N₂ mice. All mice with short REMS episode duration had the single nucleotide substitution in exon 9 of the *Nalcn* gene.



Extended Data Figure 9 | Sleep/wakefulness behaviour of *Nalcn* mutant mice. **a**, Representative 8-s EEG and EMG for wake, NREMS and REMS of *Nalcn* mutant mice **b**, Representative hypnogram of *Nalcn*^{+/+} mice (top) and *Nalcn*^{Drl/+} mice (bottom). Wake (blue), NREMS (green) and REMS (red) are indicated from ZT0 to ZT12. **c**, Enlarged hypnogram of around ZT7 showed the frequent transitions between NREMS and REMS of *Nalcn*^{Drl/+} mice. **d**, Total wake time and NREMS time of *Nalcn*^{Drl/+} mice (n = 29) and *Nalcn*^{+/+} mice (n = 25). Wake, *P* = 0.58; NREMS, *P* = 0.17,

one-way ANOVA. **e**, **f**, Circadian period length (**e**) and amplitude of circadian behaviour (**f**) in constant darkness of *Nalcn*^{Drl/+} mice (n = 6) and *Nalcn*^{+/+} mice (n = 7). *P* = 0.76 (**e**); ****P* < 0.001 (**f**), two-tailed Student's *t*-test. **g**, Total REMS time of *Nalcn*^{Drl/+} mice (n = 9) and *Nalcn*^{+/+} mice (n = 8). ****P* < 0.001, two-tailed Student's *t*-test. **h**, EEG power spectra of *Nalcn*^{Drl/+} mice (n = 29) and *Nalcn*^{+/+} mice (n = 25). ****P* < 0.001, one-way ANOVA followed by Tukey's test. Data are mean ± s.e.m.



Extended Data Figure 10 | Increased conductance of NALCN(DRL). **a–c**, *Nalcn* mRNA is expressed in the ventrolateral periaqueductal grey mater (vlPAG) and deep mesencephalic nucleus (DpMe) of the upper pons (**a**), the lateral dorsal tegmental nucleus (LDT) and sublateral dorsal nucleus (SLD) of the lower pons (**b**), and the lateral paragigantocellular nucleus (LPGi) of the medulla (**c**). AQ, aqueduct; dscp, decussation of superior cerebellar peduncle; IO, inferior olive; scp, superior cerebellar peduncle. Scale bars, 500 μ m. **d**, Representative traces of membrane currents in response to ramp pulses ($V_h = 0$ mV, from -100 mV to $+100$ mV

in 1 s; lower) recorded from HEK293T cells cotransfected with UNC80, SRC(Tyr529Phe), and NALCN-GFP (top) or NALCN(DRL)-GFP (middle). The traces are averaged from three trials. The transient capacitance currents are also recorded. **e**, Mean current density in response to ramp pulses (NALCN, $n = 5$, black line; NALCN(DRL), $n = 7$, purple line). The data from NALCN are also shown on an expanded scale (bottom right). **f**, The charge transfer of NALCN(DRL)-transfected cells was larger than that of NALCN(WT)-transfected cells. $**P < 0.01$, Mann-Whitney *U* test. The recording data are same as in **e**. Data are mean \pm s.e.m.

CRISPR/Cas9 β -globin gene targeting in human haematopoietic stem cells

Daniel P. Dever^{1*}, Rasmus O. Bak^{1*}, Andreas Reinisch², Joab Camarena¹, Gabriel Washington¹, Carmencita E. Nicolas¹, Mara Pavel-Dinu¹, Nivi Saxena¹, Alec B. Wilkens¹, Sruthi Mantri¹, Nobuko Uchida^{3†}, Ayal Hendel¹, Anupama Narla⁴, Ravindra Majeti², Kenneth I. Weinberg¹ & Matthew H. Porteus¹

The β -haemoglobinopathies, such as sickle cell disease and β -thalassaemia, are caused by mutations in the β -globin (*HBB*) gene and affect millions of people worldwide. *Ex vivo* gene correction in patient-derived haematopoietic stem cells followed by autologous transplantation could be used to cure β -haemoglobinopathies. Here we present a CRISPR/Cas9 gene-editing system that combines Cas9 ribonucleoproteins and adeno-associated viral vector delivery of a homologous donor to achieve homologous recombination at the *HBB* gene in haematopoietic stem cells. Notably, we devise an enrichment model to purify a population of haematopoietic stem and progenitor cells with more than 90% targeted integration. We also show efficient correction of the Glu6Val mutation responsible for sickle cell disease by using patient-derived stem and progenitor cells that, after differentiation into erythrocytes, express adult β -globin (HbA) messenger RNA, which confirms intact transcriptional regulation of edited *HBB* alleles. Collectively, these preclinical studies outline a CRISPR-based methodology for targeting haematopoietic stem cells by homologous recombination at the *HBB* locus to advance the development of next-generation therapies for β -haemoglobinopathies.

Allogeneic haematopoietic stem-cell transplantation demonstrates that transplantation of haematopoietic stem cells (HSCs) with only a single wild-type *HBB* gene can cure the β -haemoglobinopathies. However, this transplantation technique is limited because of graft-versus-host disease and a lack of immunologically matched donors. An alternative to using allogeneic HSCs to cure the β -haemoglobinopathies is to use homologous recombination to modify the *HBB* gene directly in autologous HSCs^{1,2}. In 1985, Smithies and colleagues³ were able to modify the human *HBB* gene by homologous recombination in a human embryonic carcinoma cell line, albeit at an extremely low frequency (10^{-6}). The subsequent discoveries that a site-specific DNA double-strand break (DSB) could stimulate homologous-recombination-mediated correction of a reporter gene and that engineered nucleases could be used to induce this DSB, formed the foundation of using homologous-recombination-mediated genome editing using engineered nucleases to modify the *HBB* gene directly^{4,5}. The ease of engineering as well as the robust activity of the CRISPR/Cas9 RNA-guided endonuclease system makes it a promising tool to apply to the continuing challenge of developing effective and safe homologous-recombination-mediated genome editing to cure β -haemoglobinopathies^{6,7}.

The CRISPR/Cas9 complex consists of the Cas9 endonuclease and a 100-nucleotide single-guide RNA (sgRNA). Target identification relies first on recognition of a 3-base-pair protospacer adjacent motif (PAM) and then on hybridization between a 20-nucleotide stretch of the sgRNA and the DNA target site, which triggers Cas9 to cleave both DNA strands⁸. DSB formation activates two highly conserved repair mechanisms: canonical non-homologous end-joining (NHEJ) and homologous recombination⁹. Through iterative cycles of break and NHEJ repair, insertions and/or deletions (INDELs) can be

created at the site of the break. By contrast, genome editing by homologous recombination requires the delivery of a DNA donor molecule to serve as a homologous template, which the cellular recombination machinery uses to repair the break by a 'copy and paste' method¹⁰. For gene-editing purposes, the homologous recombination pathway can be exploited to make precise nucleotide changes in the genome⁴. One of the key features of precise genome editing, in contrast to viral-vector-based gene transfer methods, is that endogenous promoters, regulatory elements, and enhancers can be preserved to mediate spatiotemporal gene expression^{1,11–13}. The CRISPR/Cas9 system is highly effective at stimulating DSBs in primary human haematopoietic stem and progenitor cells (HSPCs) when the sgRNA is synthesized with chemical modifications, and then electroporated into cells¹⁴.

HSCs have the ability to repopulate an entire haematopoietic system¹⁵, and several genetic^{16–18} and acquired¹⁹ diseases of the blood could potentially be cured by genome editing of HSCs. Recent studies have demonstrated efficient targeted integration in HSPCs by combining zinc-finger nuclease (ZFN) expression with exogenous homologous recombination donors delivered via single-stranded oligonucleotides²⁰, integrase-defective lentiviral vectors²¹, or recombinant adeno-associated viral vectors of serotype 6 (rAAV6)^{22,23}. While showing very positive results *in vitro*, collectively, these studies also suggested that targeting HSCs by homologous recombination at disease-causing loci is difficult in clinically relevant HSPCs.

In this study, we achieve efficient homologous-recombination-mediated editing frequencies at the *HBB* locus in CD34⁺ HSPCs derived from mobilized peripheral blood using Cas9 ribonucleoproteins combined with rAAV6 homologous donor delivery. In brief, we demonstrate: (1) Cas9- and rAAV6-mediated *HBB* targeting in HSCs characterized by the identification of modified human cells in

¹Department of Pediatrics, Stanford University, Stanford, California 94305, USA. ²Department of Medicine, Division of Hematology, Cancer Institute, and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California 94305, USA. ³Stem Cells, Inc. 7707 Gateway Blvd., Suite 140, Newark, California 94560, USA. ⁴Division of Hematology/Oncology, Department of Pediatrics, Stanford University School of Medicine, Stanford, California 94035, USA. [†]Present address: Institute of Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California 94305, USA.

*These authors contributed equally to this work.

secondary transplants of immunodeficient mice; (2) efficient correction of the sickle cell disease (SCD)-causing Glu6Val mutation in HSPCs derived from several patients with SCD; and (3) development of a purification scheme using either fluorescence-activated cell sorting (FACS) or magnetic bead enrichment to create HSPC populations in which more than 85% of the cells have been modified by homologous-recombination-mediated targeted integration. This purification can be performed early in the manufacturing process when HSCs are still preserved, and may prove valuable in a clinical setting for removing untargeted HSPCs that will be in competition with homologous-recombination-edited HSPCs for engraftment and re-population after transplantation.

CRISPR/Cas9-mediated *HBB* gene editing in HSPCs

We confirmed high transduction of HSPCs using a self-complementary AAV6 (scAAV6) with an SFFV-GFP expression cassette²⁴ (Extended Data Fig. 1a). *HBB*-specific single-stranded AAV6 (ssAAV6) vectors were then produced containing SFFV-GFP flanked by arms homologous to *HBB* (Fig. 1a). To achieve gene editing at *HBB*, we used two different CRISPR platforms, which we have previously shown to be

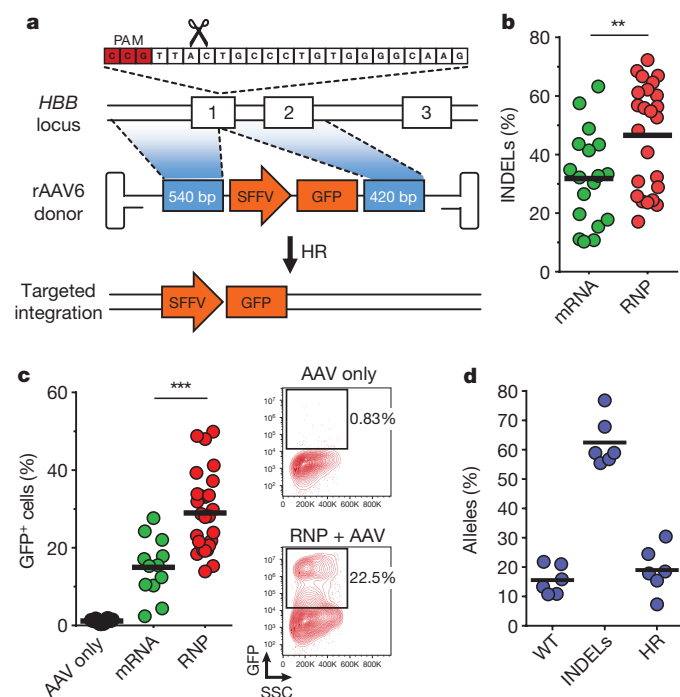


Figure 1 | CRISPR/Cas9 and rAAV6-mediated targeted integration at the *HBB* locus in human CD34⁺ HSPCs. **a**, Schematic of targeted genome editing at the *HBB* locus using CRISPR/Cas9 and rAAV6. Site-specific DSBs are created by Cas9 (scissors) mainly between nucleotides 17 and 18 of the 20-bp target site, which is followed by the 5'-NGG-3' PAM sequence (red). A DSB stimulates homologous recombination (HR) using the rAAV6 homologous donor as a repair template. White boxes, *HBB* exons; blue boxes, homology arms; orange boxes, SFFV-GFP-polyA expression cassette. **b**, HSPCs were electroporated with mRNA or RNP CRISPR system, and INDELs were analysed via TIDE software (n = number of data points within each group, all from different mobilized peripheral blood or cord blood donors; $**P < 0.01$, unpaired t -test. Bars represent means). **c**, HSPCs electroporated as above and transduced with *HBB*-specific rAAV6s were analysed by flow cytometry 18–21 days after electroporation when GFP levels were found to be constant. Left, percentage of GFP⁺ HSPCs. Right, representative FACS plots (n = number of data points within each group, all from different donors; $***P < 0.001$, unpaired t -test. Bars represent means). **d**, HSPCs were treated as above but targeted with a rAAV6 Glu6Val donor. Frequencies of allele types were quantified by sequencing of a total of 600 clones from TOPO-cloned in-out PCRs (n = 6, all from different cord blood or bone marrow donors. Bars represent means). WT, wild type.

highly active in primary cells¹⁴. Both platforms use sgRNAs chemically modified at both termini with 2'-O-methyl-3'-phosphorothioate, and are delivered either in conjunction with Cas9 mRNA or as a ribonucleoprotein (RNP) complex. Both platforms yielded high INDEL frequencies when electroporated into HSPCs, with the RNP showing higher activity (Fig. 1b). By supplying ssAAV6 *HBB* donors after electroporation of Cas9 RNP, we achieved stable green fluorescent protein (GFP) expression in an average of 29% of HSPCs (Fig. 1c). We observed lower efficiencies using the mRNA platform (15%) (Fig. 1c). Cytotoxicity and off-target cleavage activity¹⁴ was significantly decreased using the RNP system (Extended Data Fig. 1b–d).

Because AAV genomes can be captured at the site of an off-target DSB via NHEJ^{22,25,26}, we performed experiments mismatching a nuclease with a non-homologous donor to see whether this occurs with our methodology. While approximately 20% of cells that received matched *HBB* nuclease and *HBB* donor maintained GFP expression after 18 days in culture, *IL2RG* nuclease combined with *HBB* donor resulted in 0.8% GFP⁺ cells (Extended Data Fig. 1e–g), which was not significantly higher than the 0.5% GFP⁺ cells observed when using the *HBB* AAV donor alone. These results demonstrate that capture of the *HBB* donor is an infrequent event using this system in human HSPCs. Furthermore, the observed GFP expression in 0.5% of HSPCs with the AAV donor alone suggests that random integration of rAAV6 is limited in human HSPCs. In fact, previous reports have shown AAV-mediated targeted integration without a DSB^{27,28}, and thus, these GFP⁺ cells may be the result of on-target events.

SCD is caused by a single nucleotide mutation (A to T), which changes an amino acid (Glu6Val) at codon 6 of the *HBB* gene²⁹. We created a 4.5-kb rAAV6 donor template that would introduce the Glu6Val mutation along with six other silent single nucleotide polymorphisms (SNPs) to prevent Cas9 re-cutting of homologous recombination alleles (Extended Data Fig. 2a–c). Using this rAAV6 donor with Cas9 RNP delivery, we measured an average allelic modification frequency of 19% in six different HSPC donors (Fig. 1d). These results confirm that the combined use of CRISPR with rAAV6 can precisely change the nucleotide at the position of the mutation that causes SCD.

Early enrichment of *HBB*-targeted HSPCs

Because HSCs differentiate and progressively lose their long-term repopulating capacity after culturing, the identification of gene-edited HSPCs early in the manufacturing process would be of great use. In experiments using a GFP-expressing rAAV6 donor we observed that while HSPCs receiving only the rAAV6 donor expressed low levels of GFP, HSPCs that also received Cas9 RNP generated a population at day 4 after electroporation that expressed much higher levels of GFP (Fig. 2a, left). We proposed that this GFP^{high} population was enriched for *HBB*-targeted cells. We therefore sorted and cultured the GFP^{high} population as well as the GFP^{low} and GFP[−] populations. While sorted GFP^{low} and GFP[−] populations were, respectively, approximately 25% and 1% GFP⁺ after 15–20 days in culture, the GFP^{high} population was more than 95% GFP⁺, suggesting that this population was indeed *HBB*-targeted (Fig. 2a, right). Linear regression showed that the percentage of GFP^{high}-expressing HSPCs at day 4 after electroporation strongly correlates with the total percentage of GFP⁺ cells at day 18 (Extended Data Fig. 3).

To confirm that the GFP^{high} population was enriched for on-target integration, we used 'in-out PCR' (one primer binding the *HBB* locus outside the region of the homology arm and the other binding the integrated insert) to determine integration frequencies and allelic distribution in methylcellulose clones derived from the GFP^{high} population (95 clones). A total of 92% of clones had a targeted integration, with 38% containing biallelic integrations (Fig. 2b, Extended Data Fig. 4a–c and Supplementary Fig. 1a). This assay generates colonies from progenitor cells and the biallelic integration frequency could be different in HSCs. Nonetheless, these data show that the log-fold

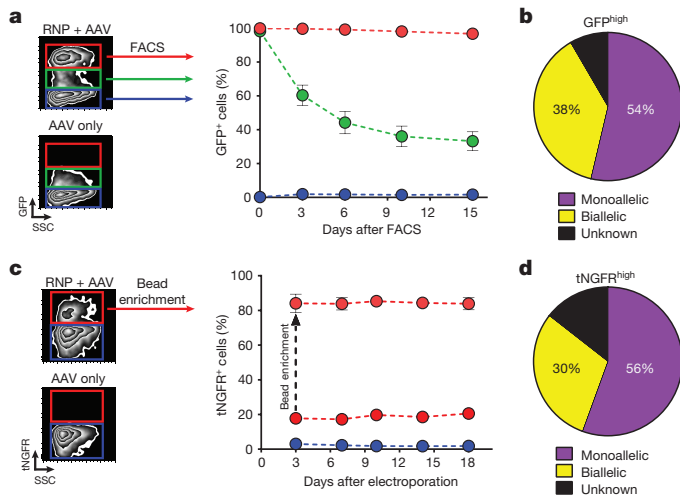


Figure 2 | Enrichment of *HBB*-targeted HSPCs using FACS and magnetic bead-based technologies. **a**, Left, representative FACS plots highlight the GFP^{high} population (red) generated by the addition of Cas9 RNP. Right, *HBB*-targeted HSPCs from GFP^{high} (red), GFP^{low} (green) and GFP⁻ (blue) fractions were sorted and monitored for GFP expression. Error bars represent s.e.m. ($n = 11$, all from unique mobilized peripheral blood or cord blood donors). **b**, PCR was performed on methylcellulose colonies from GFP^{high} HSPCs to detect targeted integration at the 3' end. **c**, Left, representative FACS plots highlight the tNGFR^{high} population (red) generated by the addition of Cas9 RNP. Right, tNGFR^{high} (red) HSPCs were enriched using anti-CD271 (also known as LNGFR) magnetic microbeads and cultured for 18 days while monitoring tNGFR expression. Error bars represent s.e.m. ($n = 5$, all from unique cord blood donors). **d**, PCR was performed on tNGFR^{high}-derived methylcellulose colonies to detect targeted integrations at the 5' end.

transgene expression shift after rAAV6 and RNP delivery is due to homologous recombination at the intended locus, and that the shift allows FACS-based enrichment of *HBB*-targeted HSPCs.

Although GFP is not a clinically relevant reporter gene, the truncated nerve growth factor receptor (tNGFR), in which the cytoplasmic intracellular signalling domain is removed, could be used to enrich for targeted HSPCs. tNGFR is expressed on the cell surface, thereby allowing antibody-mediated detection of gene marking, and has already been used in human clinical trials^{30–33}. We examined whether we could enrich *HBB*-targeted HSPCs using tNGFR magnetic-bead-based separation technology. HSPCs that received RNP and the rAAV6 donor (with a tNGFR expression cassette) produced a tNGFR^{high} population that was not present in cells transduced with rAAV6 alone (Fig. 2c, left). We then enriched tNGFR^{high} cells using anti-NGFR magnetic microbeads and after 18 days in culture, on average 84% of HSPCs were tNGFR⁺ (Fig. 2c, right). We then performed in-out PCR on tNGFR^{high} methylcellulose clones to determine on-target integration frequencies, and found that 86% of clones had a targeted integration, with 30% having biallelic integrations (Fig. 2d, Extended Data Fig. 4d–f and Supplementary Fig. 1b).

We next evaluated progenitor cell capacity of GFP^{high} HSPCs using the colony-forming unit (CFU) assay. *HBB* GFP^{high} HSPCs formed all types of colonies (erythroid, granulocyte/macrophage, and multi-lineage) to a comparable extent as mock-electroporated cells (Extended Data Fig. 5a–c). We also evaluated frequencies of GFP^{high} cells in subpopulations of HSPCs^{15,34–36} by immunophenotypic analysis, and observed a significant negative correlation between targeting frequencies and the immunophenotypic primitiveness of the analysed population (Extended Data Fig. 5d–f). To confirm these findings, we used another strategy to evaluate targeting rates in the primitive HSC population. HSPCs or HSCs were sorted from fresh cord blood, then subjected to homologous recombination experiments, and we observed a 38% reduction in targeting efficiencies in the HSCs compared to the

heterogeneous HSPC population (Extended Data Fig. 5g). We next tested whether inefficient targeting of primitive cells could be due to reduced rAAV6 donor availability. HSCs and multipotent progenitors were transduced with scAAV6-SFFV-GFP, and results showed a fivefold reduction in the number of GFP⁺ HSCs and multipotent progenitors compared to the bulk CD34⁺ population (Extended Data Fig. 5h). Collectively, our data suggest that although HSCs are more difficult to target than progenitor cells, we are achieving homologous recombination frequencies above 4% and usually above 10%.

HBB targeting in long-term repopulating HSCs

The current gold standard for HSC function, defined by the capacity to self-renew and form differentiated blood cells, is *in vivo* engraftment into immunodeficient non-obese diabetic (NOD)-severe combined immunodeficiency (SCID) *Il2rg*^{-/-} (NSG) mice. We used HSPCs derived from mobilized peripheral blood for such studies because of their high clinical relevance, although these cells have been shown to have reduced engraftment capacity in NSG mice compared to HSPCs derived from fetal liver, cord blood and bone marrow^{22,37}. All transplanted mice displayed human engraftment in the bone marrow as measured by the presence of hCD45/HLA-ABC double-positive cells 16 weeks after transplant (Fig. 3a and Extended Data Fig. 6a, b). While we observed a decrease in human cell chimaerism for all treatment groups compared to the mock-treated group, all groups with nuclease-treated cells displayed similar chimaerism to the rAAV6-only group. We did measure a small, but not statistically significant, decrease for the RNP plus AAV GFP^{high} group compared to RNP plus AAV, which can be explained by transplantation of fewer total cells and fewer phenotypically identified long-term HSCs (Extended Data Fig. 6c). There was a significant decrease from RNP plus AAV input targeting frequencies (16% in the CD34⁺ mobilized peripheral blood HSPCs) compared to the percentage of GFP⁺ cells in the bone marrow at week 16 after transplantation (3.5%) (Fig. 3c). This decrease is consistent with previous publications, and immunophenotyping of input cells did in fact show an average of 4% targeting in the CD34⁺ CD38⁻ CD90⁺ CD45RA⁻ population (Extended Data Fig. 5f). Despite these observed reductions *in vivo*, our median rates of *HBB* gene targeting in human cells in the bone marrow (3.5%) seem to be higher than what was found by others, particularly Hoban *et al.* and Genovese *et al.*, in which most mice appeared to have less than 1% gene modification after transplant using ZFNs targeting *HBB* and *IL2RG*, respectively^{20,21}. By contrast, mice transplanted with RNP plus GFP^{high} cells had a median of 90% GFP⁺ human cells at week 16 after transplant, with three mice showing more than 97% GFP⁺ human cells (Fig. 3a, c). We also observed both myeloid (CD33⁺) and lymphoid (CD19⁺) reconstitution with a median of 94% and 83% GFP⁺ cells, respectively (Fig. 3c), implicating targeting of multipotent HSCs. We detected 5% and 49% GFP⁺ human HSPCs (CD34⁺ CD10⁻) in the bone marrow of mice transplanted with RNP plus AAV and RNP plus AAV GFP^{high} cells, respectively (Extended Data Fig. 7). Multi-parameter flow cytometric analysis showed no perturbations in lineage reconstitution and no evidence of abnormal haematopoiesis, a functional assessment of the safety of the editing procedure. To determine experimentally whether we targeted *HBB* in HSCs, we performed secondary transplants for the RNP plus AAV and RNP plus AAV GFP^{high} groups. For both groups, we detected human cells in the bone marrow of secondary recipients at weeks 12–14 after transplant, with 7% and 90% GFP⁺ cells for the RNP plus AAV and RNP plus AAV GFP^{high} groups, respectively (Fig. 3d). More importantly, we confirmed *HBB* on-target integration events in human cells sorted from the bone marrow of secondary recipients from RNP plus AAV GFP^{high} experimental groups (Fig. 3e and Supplementary Fig. 2). Collectively, these data confirm that our strategy targets the *HBB* gene in human HSCs.

We next scaled up the genome-editing process to resemble a more clinically relevant HSPC starting cell number. We electroporated 80 million mobilized peripheral-blood-derived HSPCs with the *HBB*

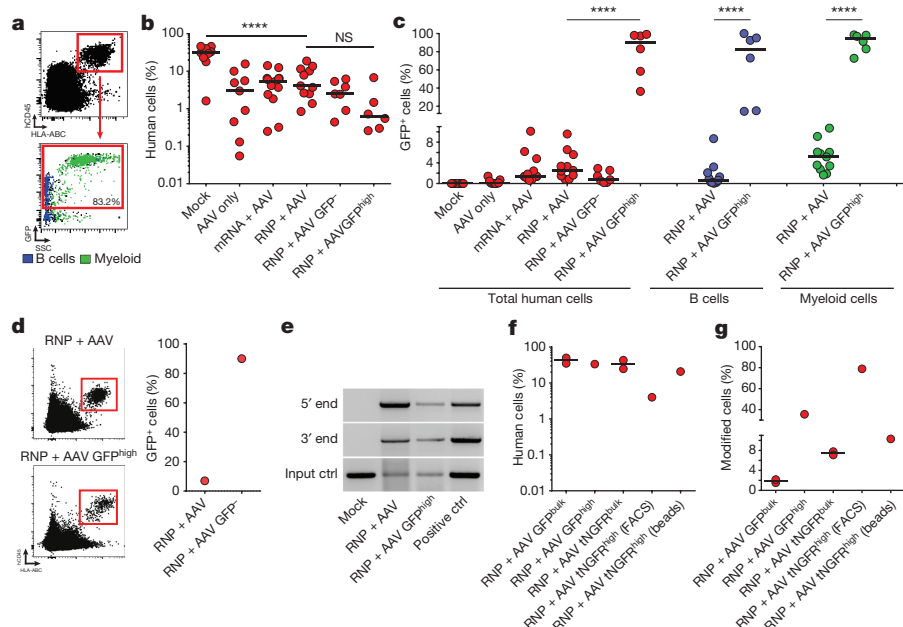


Figure 3 | *HBB* gene-targeted HSPCs display long-term and multi-lineage reconstitution in NSG mice. **a**, 16 weeks after transplantation, mouse bone marrow was analysed for human cell chimaerism and GFP expression by flow cytometry. Top, representative FACS plot from a mouse transplanted with RNP plus AAV GFP^{high} HSPCs showing engrafted human cells in the red gate. Bottom, representative FACS plot showing GFP-expressing human cells (red). CD19⁺ B cells and CD33⁺ myeloid cells are back-gated and shown in blue and green, respectively. **b**, Human engraftment in NSG mice from all experimental groups. Three different HSPC donors were used for engraftment studies (n = number of data points within group). **** P < 0.0001, NS (not significant) = P ≥ 0.05, one-way analysis of variance (ANOVA) and Tukey's multiple comparison test. Bars represent median. **c**, Percentage GFP⁺ cells in the total human population (red), CD19⁺ B cells (blue), and CD33⁺ myeloid cells (green) (n = number of data points within group). * P < 0.05, **** P < 0.0001, one-way ANOVA and Tukey's multiple comparison test for total human cells, and unpaired t -test with Welch's correction for B and myeloid cells. Bars represent median. **d**, Flow cytometric analysis of cell chimaerism

RNP system, transduced them with either SFFV-GFP or SFFV-tNGFR rAAV6, and then transplanted bulk RNP plus AAV, sorted GFP^{high} and tNGFR^{high} cells (enriched by FACS or magnetic microbeads). At 16 weeks after transplant, all mice displayed engraftment of edited human cells in the bone marrow (Fig. 3f). In this large-scale procedure, the human cell engraftment of treated cells was equivalent to 'mock' in our previous experiment (Fig. 3b). Although we observed reductions when comparing editing rates in the input cells to engrafted cells *in vivo*, the *HBB*-tNGFR mice showed a lower reduction (12% *in vitro* versus 7.5% *in vivo*) than the *HBB*-GFP mice (10% *in vitro* versus 1.9% *in vivo*), suggesting tNGFR might be a better transgene to evaluate editing of HSCs *in vivo* (Fig. 3g). Furthermore, mice transplanted with enriched targeted HSPCs displayed human cell-editing frequencies of 10–75% (three mice), with human engraftment levels ranging from 4% to 30% (Fig. 3f, g). These data suggest that our methodology can be translated to perform large-scale genome editing in HSCs at the *HBB* locus.

***HBB* gene correction in SCD HSPCs**

We next tested our methodology to correct the disease-causing Glu6Val mutation in CD34⁺ HSPCs derived from patients with SCD. We first confirmed high frequencies of INDELs (Fig. 4a) and homologous recombination using an SFFV-GFP donor (Fig. 4b) at the *HBB* locus in SCD HSPCs. We then produced a therapeutic rAAV6 donor (corrective SNP donor) designed to revert the Glu6Val mutation, while also introducing silent mutations to prevent Cas9 re-cutting and premature cross

and GFP expression 12–14 weeks after secondary transplantation. Left, representative FACS plot from a secondary mouse transplanted with RNP plus AAV (top) or RNP plus AAV GFP^{high} (bottom) cells showing engrafted human cells in the red gate. Right, percentage GFP⁺ human cells in the bone marrow of secondary recipients. **e**, Gel images of in-out PCR analyses on sorted human cells from secondary recipients to analyse on-target integrations at the 5' and 3' ends. Input control (ctrl) PCR was performed on the human *CCR5* gene. Positive control is an HSPC sample targeted at *HBB* with SFFV-GFP-polyA. **f**, 80 million mobilized peripheral-blood-derived CD34⁺ cells were electroporated with *HBB* RNP and transduced with *HBB* AAV6s. Bulk HSPCs or HSPCs enriched for targeting (by FACS or bead enrichment) were transplanted into the tail vein of sublethally irradiated mice. Then, 16 weeks after transplant, human cell chimaerism was analysed by flow cytometry (n = number of data points within group). **g**, Percentage GFP⁺ and tNGFR⁺ cells in the human population was analysed by flow cytometry (n = number of data points within group), bars represent median.

over (Extended Data Fig. 8a). Targeting SCD HSPCs with the corrective SNP donor reverted an average of 50% of the Glu6Val (HbS) alleles to wild-type (HbA) alleles (Fig. 4c), and analysis of methylcellulose clones showed that an average of 45% of clones had at least one HbA allele (Extended Data Fig. 8b). We next created an anti-sickling *HBB* cDNA therapeutic donor (HbAS3; ref. 38) using previously reported strategies^{12,13} of knocking in divergent cDNAs into the gene start codon to preserve endogenous promoter/enhancer function, followed by a clinically relevant promoter (EF1 α) driving tNGFR expression to allow for enrichment of edited cells (Extended Data Fig. 9a). Using this donor, we targeted an average of 11% of SCD patient-derived HSPCs (Fig. 4d) and confirmed seamless integration by in-out PCR (Extended Data Fig. 9b). Notably, we observed a tNGFR^{high} population as described previously (Fig. 2c), indicating the ability to enrich SCD-corrected HSPCs early in the culture process. We conclude that our methodology can correct the Glu6Val mutation using two different donor designs.

We next tested whether the *HBB*-edited SCD HSPCs maintained their erythroid differentiation potential by subjecting tNGFR^{high} and mock HSPCs to a 21-day *in vitro* erythroid differentiation protocol^{39,40}. Flow cytometric analyses after erythroid differentiation showed a high proportion of GPA⁺ CD45⁺ CD71⁺ CD34⁺ cells, indicating the presence of mature differentiated erythrocytes that express haemoglobin⁴¹ (Fig. 4e and Extended Data Fig. 10). To confirm that adult β -globin (HbA) or HbAS3 mRNA was transcribed from edited *HBB* alleles, we performed reverse transcription quantitative PCR

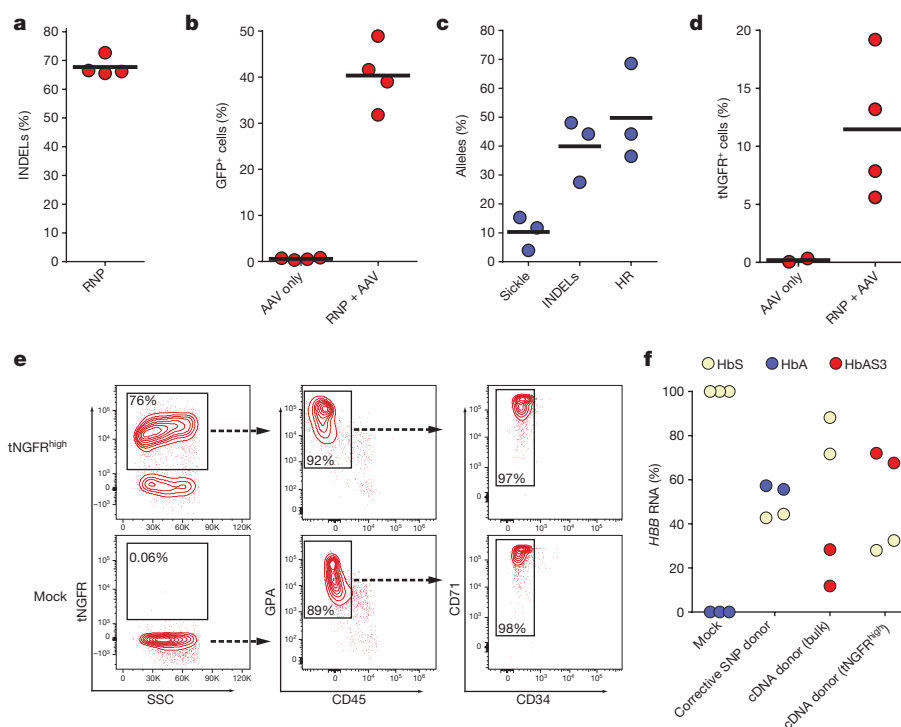


Figure 4 | Correction of the Glu6Val mutation in SCD patient-derived HSPCs. **a**, Genomic DNA from *HBB* RNP-treated SCD HSPCs was collected and INDELs were analysed via TIDE software ($n = 4$ different SCD patient donors). **b**, *HBB*-targeted SCD HSPCs were analysed for GFP expression by flow cytometry ($n = 4$ different SCD patient donors). **c**, SCD HSPCs were targeted with the rAAV6 corrective SNP donor. *HBB* allele types were analysed by sequencing TOPO-cloned PCR fragments derived from in-out PCR. Around 50–100 TOPO clones were analysed from each of three different HSPC donors ($n = 3$ different SCD patient donors). **d**, SCD HSPCs were targeted with the anti-sickling *HBB* cDNA-EF1 α

-tNGFR correction donor. Frequencies of tNGFR⁺ cells were analysed by flow cytometry ($n = 3$ different SCD patient donors). **e**, SCD mock HSPCs and sorted SCD tNGFR^{high} HSPCs were differentiated into erythrocytes *in vitro*. Representative FACS plots from day 21 of differentiation show cell surface markers associated with erythrocytes (GPA⁺ CD45[−] CD71⁺ CD34[−]). **f**, HbS, HbA and HbAS3 mRNA expression was quantified by RT-qPCR in erythrocytes differentiated from *HBB*-edited or mock SCD HSPCs. All mRNA transcript levels were normalized to the *RPLP0* input control ($n = 2–3$ different SCD patient donors). All bars represent means.

(RT-qPCR) on erythrocytes differentiated from edited SCD HSPCs. Erythrocytes edited with the corrective SNP donor expressed 56% HbA mRNA out of total β -globin mRNA, whereas erythrocytes edited with the cDNA donor (bulk) expressed 20% HbAS3 mRNA (Fig. 4f). Thus the percentage of HbAS3 mRNA (20%) matched or exceeded the percentage of cells modified by the tNGFR cassette (11%), suggesting functional expression of the AS3 cDNA from the endogenous *HBB* promoter. Notably, erythrocytes differentiated from enriched tNGFR^{high} HSPCs expressed 70% HbAS3 mRNA, confirming an enrichment of functionally corrected HSPCs.

Discussion

While our data support the idea that HSCs are more resistant to homologous recombination-mediated editing, they also show that it is possible to edit these cells at reasonable frequencies²¹. Enrichment of targeted cells resulted in the removal of most HSCs, leading to an overall eightfold decrease in the total number of HSCs in the transplanted enriched population (Extended Data Fig. 6c). Even though we transplanted eightfold fewer total HSCs in the enriched population than the non-enriched population (GFP^{high} versus RNP plus AAV), we found that absolute numbers of edited cells in the bone marrow of the mice from the enriched group was on average fivefold higher 16 weeks after transplant (Extended Data Fig. 6c). Thus, our enrichment strategy not only yields higher frequencies of modified cells in the transplanted mice, but also the absolute number of modified human cells in the mice was higher and can thus ameliorate the problem of inefficient HSC targeting. Recent advances in *ex vivo* HSC expansion protocols and identification of small molecule drugs, such as UM171 (ref. 42) that expands HSCs, might be combined with our strategy to generate a large and highly enriched population of genome-edited HSCs. Future studies

will help to determine whether HSC expansion would be required for clinical translation of our enrichment model.

Our methodology sets the framework for CRISPR-mediated *HBB* targeting in HSCs that has the power to be translated to the clinic. Although GFP is an unsuitable marker for gene therapy, our enrichment protocol using tNGFR (Figs 2c, 3f, g) (or other similar signalling-inert cell surface markers) represents a strategy for the next generation of β -haemoglobinopathy therapies that are based on gene editing. These studies show that this methodology can enrich corrected SCD patient-derived HSPCs that can differentiate into erythrocytes that express *HBB* anti-sickling mRNA from the endogenous *HBB* promoter. This tNGFR selection strategy has the potential advantage over chemoselection strategies because it avoids exposing edited cells and patients to potentially toxic chemotherapy¹³. The strategy of knocking in a *HBB* cDNA along with a selectable marker to enrich for modified cells would be applicable to both SCD and almost all forms of β -thalassaemia. Furthermore, because we can efficiently scale up the genome-editing process to clinically relevant starting numbers, future studies will address the development of a current good manufacturing practice (cGMP)-compatible process for editing the *HBB* locus in HSPCs.

In conclusion, we believe that the presented methodology lays the foundation for CRISPR/Cas9-mediated genome editing therapies not only for the β -haemoglobinopathies, but also for a range of other haematological diseases, and generally advances HSC-based cell and gene therapies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 May; accepted 29 September 2016.

Published online 7 November 2016.

1. Naldini, L. *Ex vivo* gene transfer and correction for cell-based therapies. *Nat. Rev. Genet.* **12**, 301–315 (2011).
2. Porteus, M. Genome editing: a new approach to human therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **56**, 163–190 (2016).
3. Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal β -globin locus by homologous recombination. *Nature* **317**, 230–234 (1985).
4. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
5. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl Acad. Sci. USA* **91**, 6064–6068 (1994).
6. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
7. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
8. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
9. Kass, E. M. & Jasin, M. Collaboration and competition between DNA double-strand break repair pathways. *FEBS Lett.* **584**, 3703–3708 (2010).
10. Porteus, M. H. Towards a new era in medicine: therapeutic genome editing. *Genome Biol.* **16**, 286 (2015).
11. Woods, N. B., Bottero, V., Schmidt, M., von Kalle, C. & Verma, I. M. Gene therapy: therapeutic gene causing lymphoma. *Nature* **440**, 1123 (2006).
12. Hubbard, N. *et al.* Targeted gene editing restores regulated CD40L expression and function in X-HIGM T cells. *Blood* **127**, 2513–2522 (2016).
13. Voit, R. A., Hendel, A., Pruett-Miller, S. M. & Porteus, M. H. Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res.* **42**, 1365–1378 (2014).
14. Hendel, A. *et al.* Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* **33**, 985–989 (2015).
15. Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A. M. & Peault, B. Isolation of a candidate human hematopoietic stem-cell population. *Proc. Natl Acad. Sci. USA* **89**, 2804–2808 (1992).
16. Mukherjee, S. & Thrasher, A. J. Gene therapy for PIDs: progress, pitfalls and prospects. *Gene* **525**, 174–181 (2013).
17. Cavazzana-Calvo, M. *et al.* Transfusion independence and HMGA2 activation after gene therapy of human β -thalassaemia. *Nature* **467**, 318–322 (2010).
18. Naldini, L. Gene therapy returns to centre stage. *Nature* **526**, 351–360 (2015).
19. Jenq, R. R. & van den Brink, M. R. Allogeneic hematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. *Nat. Rev. Cancer* **10**, 213–221 (2010).
20. Hoban, M. D. *et al.* Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood* **125**, 2597–2604 (2015).
21. Genovese, P. *et al.* Targeted genome editing in human repopulating hematopoietic stem cells. *Nature* **510**, 235–240 (2014).
22. Wang, J. *et al.* Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat. Biotechnol.* **33**, 1256–1263 (2015).
23. De Ravin, S. S. *et al.* Targeted gene addition in human CD34⁺ hematopoietic cells for correction of X-linked chronic granulomatous disease. *Nat. Biotechnol.* **34**, 424–429 (2016).
24. Sather, B. D. *et al.* Efficient modification of *CCR5* in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci. Transl. Med.* **7**, 307ra156 (2015).
25. Wang, J. *et al.* Highly efficient homology-driven genome editing in human T cells by combining zinc-finger nuclease mRNA and AAV6 donor delivery. *Nucleic Acids Res.* **44**, e30 (2016).
26. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* **36**, 767–773 (2004).
27. Russell, D. W. & Hirata, R. K. Human gene targeting by viral vectors. *Nat. Genet.* **18**, 325–330 (1998).
28. Barzel, A. *et al.* Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* **517**, 360–364 (2015).
29. Hoban, M. D., Orkin, S. H. & Bauer, D. E. Genetic treatment of a molecular disorder: gene therapy approaches to sickle cell disease. *Blood* **127**, 839–848 (2016).
30. Bonini, C. *et al.* HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukemia. *Science* **276**, 1719–1724 (1997).
31. Ciceri, F. *et al.* Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haemopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I-II study. *Lancet Oncol.* **10**, 489–500 (2009).
32. Oliveira, G. *et al.* Tracking genetically engineered lymphocytes long-term reveals the dynamics of T cell immunological memory. *Sci. Transl. Med.* **7**, 317ra198 (2015).
33. Bonini, C. *et al.* Safety of retroviral gene marking with a truncated NGF receptor. *Nat. Med.* **9**, 367–369 (2003).
34. Seita, J. & Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 640–653 (2010).
35. Majeti, R., Park, C. Y. & Weissman, I. L. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* **1**, 635–645 (2007).
36. Doulatov, S., Notta, F., Lauretti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120–136 (2012).
37. Gu, A. *et al.* Engraftment and lineage potential of adult hematopoietic stem and progenitor cells is compromised following short-term culture in the presence of an aryl hydrocarbon receptor antagonist. *Hum. Gene Ther. Methods* **25**, 221–231 (2014).
38. Levasseur, D. N. *et al.* A recombinant human hemoglobin with anti-sickling properties greater than fetal hemoglobin. *J. Biol. Chem.* **279**, 27518–27524 (2004).
39. Dulmovits, B. M. *et al.* Pomalidomide reverses γ -globin silencing through the transcriptional reprogramming of adult hematopoietic progenitors. *Blood* **127**, 1481–1492 (2016).
40. Hu, J. *et al.* Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis *in vivo*. *Blood* **121**, 3246–3253 (2013).
41. Romero, Z. *et al.* β -globin gene transfer to human bone marrow for sickle cell disease. *J. Clin. Invest.* **123**, 3317–3330 (2013).
42. Fares, I. *et al.* Cord blood expansion. Pyrimidinole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science* **345**, 1509–1512 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements D.P.D. was supported by a Stanford Child Health Research Institute (CHRI) Grant and Postdoctoral Award. R.O.B. was supported by an Individual Postdoctoral grant (DFF-1333-00106B) and a Sapere Aude, Research Talent grant (DFF-1331-00735B), both from the Danish Council for Independent Research, Medical Sciences. M.H.P. acknowledges the support of the Amos Carter Foundation, the Laurie Kraus Jacob Faculty Scholar Award in Pediatric Translational Research and NIH grant support PN2EY018244, R01-AI097320 and R01-AI120766. We thank D. Russell for the pDGM6 plasmid, H.-P. Kiem for scAAV6, G. de Alencastro and M. Kay for help with AAV production, the Binns Program for Cord Blood Research at Stanford University for cord-blood-derived CD34⁺ HSPCs. We also thank Lonza (A. Toell and G. Alberts) for donating the LV unit for performing large-scale genome-editing studies. We further thank members of the Porteus laboratory, D. DiGiusto and M. G. Roncarolo for input, comments and discussion.

Author Contributions D.P.D. and R.O.B. contributed equally to this work as well as performed and designed most of the experiments. D.P.D., R.O.B., A.R. and C.E.N. designed, executed and analysed engraftment studies. J.C. performed some *HBB* homologous recombination, NHEJ and tNGFR experiments. G.W. performed *in vitro* erythrocyte differentiation experiments. M.P.-D. assisted with large-scale electroporation experiments. N.S. ran and analysed HSPC immunophenotyping experiments. A.B.W. scored and collected methylcellulose clones. N.U. sorted freshly isolated HSCs before targeting. S.M. purified CD34⁺ HSPCs from peripheral blood of patients with SCD. A.N. was the attending physician who oversaw transfusions of patients with SCD. A.H., R.M. and K.I.W. contributed to experimental design and data interpretation. M.H.P. directed the research and participated in the design and interpretation of the experiments and the writing of the manuscript. D.P.D. and R.O.B. wrote the manuscript with help from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H.P. (mporteus@stanford.edu).

Reviewer Information *Nature* thanks B. Ebert, L. Naldini and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

AAV vector production. AAV vector plasmids were cloned in the pAAV-MCS plasmid (Agilent Technologies) containing inverted terminal repeats from AAV serotype 2. The *HBB* rAAV6 GFP and tNGFR donor contained promoter, MaxGFP or tNGFR, and BGH polyA. The left and right homology arms for the GFP and tNGFR *HBB* donors were 540 bp and 420 bp, respectively. The Glu6Val rAAV6 donor contained 2.2 kb of sequence homologous to the sequence upstream of Glu6Val. The nucleotide changes are depicted in Extended Data Fig. 2. Immediately downstream of the last nucleotide change was 2.2 kb of homologous *HBB* sequence. *HBB* cDNA contained same homology arms as GFP and tNGFR donors above except the left homology arm was shortened to end at the sickle mutation. Sequence of full *HBB* cDNA is depicted in (Extended Data Fig. 9b). The sickle corrective donor used in the SCD-derived HSPCs in Fig. 4 had a total of 2.4 kb sequence homology to *HBB* with the SNPs shown in Extended Data Fig. 8a in the centre. scAAV6 carrying the SFFV promoter driving GFP was provided by H.-P. Kiem. AAV6 vectors were produced as described with a few modifications⁴³. In brief, 293FT cells (Life Technologies) were seeded at 13×10^6 cells per dish in ten 15-cm dishes one day before transfection. One 15-cm dish was transfected using standard PEI transfection with 6 μ g ITR-containing plasmid and 22 μ g pDGM6 (a gift from D. Russell), which contains the AAV6 cap genes, AAV2 rep genes, and adenovirus helper genes. Cells were incubated for 72 h until collection of AAV6 from cells by three freeze–thaw cycles followed by a 45 min incubation with TurboNuclease at 250 U ml⁻¹ (Abnova). AAV vectors were purified on an iodixanol density gradient by ultracentrifugation at 237,000g for 2 h at 18 °C. AAV vectors were extracted at the 60–40% iodixanol interface and dialysed three times in PBS with 5% sorbitol in the last dialysis using a 10K MWCO Slide-A-Lyzer G2 Dialysis Cassette (Thermo Fisher Scientific). Vectors were added pluronic acid to a final concentration of 0.001%, aliquoted, and stored at –80 °C until use. AAV6 vectors were titred using quantitative PCR to measure number of vector genomes as described previously⁴⁴.

CD34⁺ haematopoietic stem and progenitor cells. Frozen CD34⁺ HSPCs derived from bone marrow or mobilized peripheral blood were purchased from AllCells and thawed according to manufacturer's instructions. CD34⁺ HSPCs from cord blood were either purchased frozen from AllCells or acquired from donors under informed consent via the Binns Program for Cord Blood Research at Stanford University and used fresh without freezing. CD34⁺ HSPCs from patients with SCD were purified within 24 h of the scheduled apheresis. For volume reduction via induced rouleaux formation, whole blood was added 6% Hetastarch in 0.9% sodium chloride injection (Hospira, Inc.) in a proportion of 5:1 (v/v). Following a 60–90-min incubation at room temperature, the top layer, enriched for HSPCs and mature leukocytes, was carefully isolated with minimal disruption of the underlying fraction. Cells were pelleted, combined, and resuspended in a volume of PBS with 2 mM EDTA and 0.5% BGS directly proportional to the fraction of residual erythrocytes—typically 200–400 ml. Mononuclear cells (MNCs) were obtained by density gradient separation using Ficoll and CD34⁺ HSPCs were purified using the CD34⁺ Microbead Kit Ultrapur (Miltenyi Biotec) according to manufacturer's protocol. Cells were cultured overnight and then stained for CD34 and CD45 using APC anti-human CD34 (clone 561; Biolegend) and BD Horizon V450 anti-human CD45 (clone HI30; BD Biosciences), and a pure population of HSPCs defined as CD34^{bright}/CD45^{dim} were obtained by cell sorting on a FACS Aria II cell sorter (BD Biosciences). All CD34⁺ HSPCs were cultured in StemSpan SFEM II (StemCell Technologies) supplemented with SCF (100 ng ml⁻¹), TPO (100 ng ml⁻¹), Flt3 ligand (100 ng ml⁻¹), IL-6 (100 ng ml⁻¹), and StemRegenin1 (0.75 mM). Cells were cultured at 37 °C, 5% CO₂ and 5% O₂.

Electroporation and transduction of cells. The *HBB* and *IL2RG* synthetic sgRNAs used were purchased from TriLink BioTechnologies with chemically modified nucleotides at the three terminal positions at both the 5' and 3' ends. Modified nucleotides contained 2'-O-methyl-3'-phosphorothioate and the sgRNAs were HPLC-purified. The genomic sgRNA target sequences, with PAM in bold, are: *HBB*: 5'-CTTGCCCCACAGGGCAGTAACGG-3' (refs 45, 46); *IL2RG*: 5'-TGGAATGATGGCTTCAACATGG-3'. Cas9 mRNA containing 5-methylcytidine and pseudouridine was purchased from TriLink BioTechnologies. Cas9 protein was purchased from Life Technologies. Cas9 RNP was made by incubating protein with sgRNA at a molar ratio of 1:2.5 at 25 °C for 10 min immediately before electroporation. CD34⁺ HSPCs were electroporated 1–2 days after thawing or isolation. CD34⁺ HSPCs were electroporated using the Lonza Nucleofector 2b (program U-014) and the Human T Cell Nucleofection Kit (VPA-1002, Lonza) as we have found this combination to be superior in optimization studies. The following conditions were used: 5×10^6 cells ml⁻¹, 300 μ g ml⁻¹ Cas9 protein complexed with sgRNA at 1:2.5 molar ratio, or 100 μ g ml⁻¹ synthetic chemically modified sgRNA with 150 μ g ml⁻¹ Cas9 mRNA (TriLink BioTechnologies, non-HPLC purified). Following electroporation, cells were incubated for 15 min at 37 °C after which they were added AAV6 donor

vectors at an MOI (vector genomes/cell) of 50,000–100,000 and then incubated at 30 °C or 37 °C overnight (if incubated at 30 °C, plates were then transferred to 37 °C) or targeting experiments of freshly sorted HSCs (Extended Data Fig. 5g), cells were electroporated using the Lonza Nucleofector 4D (program EO-100) and the P3 Primary Cell Nucleofection Kit (V4XP-3024). For the electroporation of 80 million CD34⁺ HSPCs, the Lonza 4D-Nucleofector LV unit (program DZ-100) and P3 Primary Cell Kit were used. Subsequently, we have found no benefit to the 30 °C incubation and now perform all of our manufacturing at 37 °C.

Measuring targeted integration of fluorescent and tNGFR donors. Rates of targeted integration of GFP and tNGFR donors were measured by flow cytometry at least 18 days after electroporation. Targeted integration of a tNGFR expression cassette was measured by flow cytometry of cells stained with APC-conjugated anti-human CD271 (NGFR) antibody (BioLegend, clone: ME20.4). For sorting of GFP^{high} or tNGFR^{high} populations, cells were sorted on a FACS Aria II SORP using the LIVE/DEAD Fixable Blue Dead Cell Stain Kit (Life Technologies) to discriminate live and dead cells according to manufacturer's instructions.

Positive selection and enrichment of tNGFR⁺ HSPCs. Positive selection of targeted HSPCs was performed using the CD271 (tNGFR) Microbead Kit (Miltenyi Biotec), according to the manufacturer's instructions 72 h after electroporation. In brief, tNGFR⁺ cells were magnetically labelled with CD271 Microbeads after which the cell suspension was loaded onto an equilibrated MACS column inserted in the magnetic field of a MACS separator. The columns were washed three times, and enriched cells were eluted by removing the column from the magnetic field and eluting with PBS. Enrichment was determined by flow cytometry during culture for 2–3 weeks by FACS analysis every 3 days.

Immunophenotyping of targeted HSPCs. Collected wells were stained with LIVE/DEAD Fixable Blue Dead Cell Stain (Life Technologies) and then with anti-human CD34 PE-Cy7 (581, BioLegend), CD38 Alexa Fluor 647 (AT1, Santa Cruz Biotechnologies), CD45RA BV 421 (HI100, BD Biosciences), and CD90 BV605 (5E10, BioLegend) and analysed by flow cytometry. For sorting of CD34⁺ or CD34⁺ CD38⁻ CD90⁺ cells, cord-blood-derived CD34⁺ HSPCs were stained directly after isolation from blood with anti-human CD34 FITC (8G12, BD Biosciences), CD90 PE (5E10, BD Biosciences), CD38 APC (HIT2, BD Bioscience), and cells were sorted on a FACS Aria II (BD Bioscience), cultured overnight, and then electroporated with *HBB* RNP and transduced with *HBB* GFP rAAV6 using our optimized parameters.

Measuring targeted integration of the E6V donor. For assessing the allele modification frequencies in samples with targeted integration of the Glu6Val rAAV6 donor, PCR amplicons spanning the targeted region (see Extended Data Fig. 2a) were created using one primer outside the donor homology arm and one inside: *HBB*_outside 5'-GGTGACAATTTCTGCCAATCAGG-3' and *HBB*_inside: 5'-GAATGGTAGCTGGATTGACTGTC-3'. The PCR product was gel-purified and re-amplified using a nested primer set (*HBB*_nested_fw: 5'-GAAGATATGCTTAGAACCGAGG-3' and *HBB*_nested_rv: 5'-CCACATGCCAGTTTCTATTGG-3') to create a 685-bp PCR amplicon (see Extended Data Fig. 2a) that was gel-purified and cloned into a TOPO plasmid using the Zero Blunt TOPO PCR Cloning Kit (Life Technologies) according to the manufacturer's protocol. TOPO reactions were transformed into XL-1 Blue competent cells, plated on kanamycin-containing agar plates, and single colonies were sequenced by McLab by rolling circle amplification followed by sequencing using the following primer: 5'-GAAGATATGCTTAGAACCGAGG-3'. For each of the six unique CD34⁺ donors used in this experiment, 100 colonies were sequenced. Additionally, 100 colonies derived from an AAV-only sample were sequenced and detected no integration events.

Measuring INDEL frequencies. INDEL frequencies were quantified using the TIDE software⁴⁷ (tracking of indels by decomposition) and sequenced PCR products obtained by PCR of genomic DNA extracted at least 4 days after electroporation as previously described¹⁴.

Methylcellulose CFU assay. The CFU assay was performed by FACS sorting of single cells into 96-well plates containing MethoCult Optimum (StemCell Technologies) 4 days after electroporation and transduction. After 12–16 days, colonies were counted and scored based on their morphological appearance in a blinded fashion.

Genotyping of methylcellulose colonies. DNA was extracted from colonies formed in methylcellulose from FACS sorting of single cells into 96-well plates. In brief, PBS was added to wells with colonies, and the contents were mixed and transferred to a U-bottomed 96-well plate. Cells were pelleted by centrifugation at 300g for 5 min followed by a wash with PBS. Finally, cells were resuspended in 25 μ l QuickExtract DNA Extraction Solution (Epicentre) and transferred to PCR plates, which were incubated at 65 °C for 10 min followed by 100 °C for 2 min. Integrated or non-integrated alleles were detected by PCR. For detecting *HBB* GFP integrations at the 3' end, two different PCRs were set up to detect

integrated (one primer in insert and one primer outside right homology arm) and non-integrated alleles (primer in each homology arm), respectively (see Extended Data Fig. 4a). *HBB*_int_fw: 5'-GTACCAGCACGCCTTCAAGACC-3', *HBB*_int_rv: 5'-GATCCTGAGACTTCCACACTGATGC-3', *HBB*_no_int_fw: 5'-GAAGATATGCTTAGAACCGAGG-3', *HBB*_no_int_rv: 5'-CCACATGCCAGTTTCTATTGG-3'. For detecting *HBB* tNGFR integrations at the 5' end, a 3-primer PCR methodology was used to detect the integrated and non-integrated allele simultaneously (see Extended Data Fig. 4d). *HBB*_outside_5'Arm_fw: 5'-GAAGATATGCTTAGAACCGAGG-3', *SFFV*_rev: 5'-ACCGCAGATATCCTGTTGG-3', *HBB*_inside_3'Arm_rev: 5'-CCACATGCCAGTTTCTATTGG-3'. Note that for the primers assessing non-integrated alleles, the Cas9 cut site is at least 90 bp away from the primer-binding sites and since CRISPR/Cas9 generally introduces INDELS of small sizes, the primer-binding sites should only very rarely be disrupted by an INDEL.

Transplantation of CD34⁺ HSPCs into NSG mice. For *in vivo* studies, 6 to 8 week-old NSG mice were purchased from the Jackson laboratory (Bar Harbour). The experimental protocol was approved by Stanford University's Administrative Panel on Laboratory Animal Care. For transplant data in Fig. 3a–c, sample sizes were not chosen to ensure adequate power to detect a pre-specified effect size. Four days after electroporation/transduction or directly after sorting, 500,000 cells (or 100,000–500,000 cells for the GFP^{high} group) were administered by tail-vein injection into the mice after sub-lethal irradiation (200 cGy) using an insulin syringe with a 27 gauge \times 0.5 inch (12.7 mm) needle. For transplant data in Fig. 3f, g, three days after electroporation, 400,000–700,000 bulk HSPCs or HSPCs enriched for targeting (FACS or bead-enrichment) were transplanted as described above. Mice were randomly assigned to each experimental group and evaluated in a blinded fashion. For secondary transplants, human cells from the RNP plus AAV group were pooled and CD34⁺ cells were selected using a CD34 bead enrichment kit (MACS CD34 MicroBead Kit UltraPure, human, Miltenyi Biotec), and finally cells were injected into the femurs of female secondary recipients (3 mice total). Because GFP^{high} mice had low engraftment, they were not CD34⁺-selected, but total mononuclear cells were filtered, pooled, and finally injected into the femur of two secondary recipients.

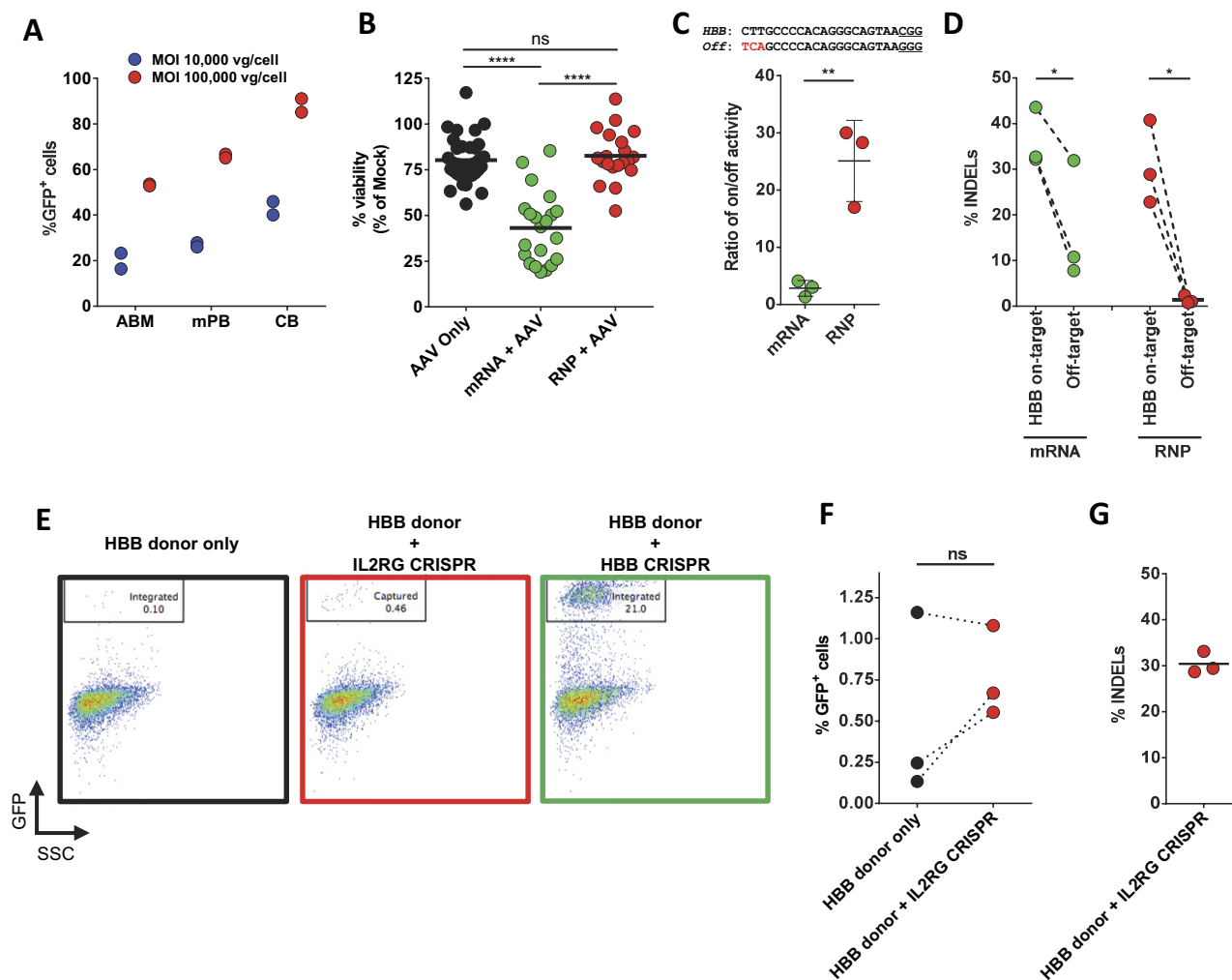
Assessment of human engraftment. At week 16 after transplantation, mice were euthanized, mouse bones (2 \times femur, 2 \times tibia, 2 \times humerus, sternum, 2 \times pelvis, spine) were collected and crushed using mortar and pestle. MNCs were enriched using Ficoll gradient centrifugation (Ficoll-Paque Plus, GE Healthcare) for 25 min at 2,000g, room temperature. Cells were blocked for nonspecific antibody binding (10% v/v, TruStain FcX, BioLegend) and stained (30 min, 4°C, dark) with monoclonal anti-human CD45 V450 (HI30, BD Biosciences), CD19 APC (HIB19, BD Biosciences), CD33 PE (WM53, BD Biosciences), HLA-ABC APC-Cy7 (W6/32, BioLegend), anti-mouse CD45.1 PE-Cy7 (A20, eBioScience), anti-mouse PE-Cy5 mTer119 (TER-119, eBioScience) antibodies. Normal multi-lineage engraftment was defined by the presence of myeloid cells (CD33⁺) and B-cells (CD19⁺) within engrafted human CD45⁺ HLA-ABC⁺ cells. Parts of the mouse bone marrow were used for CD34-enrichment (MACS CD34 MicroBead Kit UltraPure, human, Miltenyi Biotec) and the presence of human HSPCs was assessed by staining with anti-human CD34 APC (8G12, BD Biosciences), CD38 PE-Cy7 (HB7, BD Biosciences), CD10 APC-Cy7 (HI10a, BioLegend), and anti-mouse CD45.1 PE-Cy5 (A20, eBioScience) and analysed by flow cytometry. The estimation of the total number of modified human cells in the bone marrow at week 16 after transplant was calculated by multiplying the percentage engraftment with the percentage GFP⁺ cells among engrafted cells. This number was multiplied by the total number of MNCs in the bone marrow of a NSG mouse (1.1 \times 10⁸ per mouse) to give the total number of GFP⁺ human cells in the total bone marrow of the transplanted mice. The total number of MNCs in the bone marrow of a NSG mouse was calculated by counting the total number of MNCs in one femur in four NSG mice. The total number of MNCs in one mouse was then calculated assuming one femur is 6.1% of the total marrow as found previously⁴⁸.

Differentiation of CD34⁺ HSPCs into erythrocytes *in vitro*. SCD patient-derived HSPCs were cultured in three phases following targeting at 37°C and 5% CO₂ in SFEM II media according to previously established protocols^{39,40}. Media was supplemented with 100 U ml⁻¹ penicillin/streptomycin, 2 mM L-glutamine, 40 μ g ml⁻¹ lipids, 100 ng ml⁻¹ SCF, 10 ng ml⁻¹ IL-3 (PeproTech), 0.5 U ml⁻¹ erythropoietin (eBiosciences), and 200 μ g ml⁻¹ transferrin (Sigma Aldrich). In the first phase, corresponding to days 0–7 (day 0 being day 4 after electroporation), cells were cultured at 10⁵ cells ml⁻¹. In the second phase, corresponding to days 7–11, cells were maintained at 10⁵ cells ml⁻¹ and erythropoietin was increased to 3 U ml⁻¹. In the third and final phase, days 11–21, cells were cultured at 10⁶ cells ml⁻¹ with 3 U ml⁻¹ of erythropoietin and 1 mg ml⁻¹ of transferrin. Erythrocyte differentiation of edited and non-edited HSPCs was assessed by flow cytometry using the following antibodies: hCD45 V450 (HI30, BD Biosciences), CD34 FITC (8G12, BD Biosciences), CD71 PE-Cy7 (OKT9, Affymetrix), and CD235a PE (GPA) (GA-R2, BD Biosciences).

Assessment of mRNA levels in differentiated erythrocytes. RNA was extracted from 100,000–250,000 differentiated erythrocytes between days 16–21 of erythroid differentiation using the RNeasy Mini Kit (Qiagen) and was DNase-treated with RNase-Free DNase Set (Qiagen). cDNA was made from 100 ng RNA using the iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad). Levels of HbS, HbA (from corrective SNP donor), and HbA-AS3 (anti-sickling *HBB* cDNA donor) were quantified by qPCR using the following primers and FAM/ZEN/IBFQ-labelled hydrolysis probes purchased as custom-designed PrimeTime qPCR Assays from IDT: HbS primer (fw): 5'-TCACTAGCAACCTCAAACAGAC-3', HbS primer (rv): 5'-ATCCACGTTACCTTGCC-3', HbS probe: 5'-TAACG GCAGACTTCTCCACAGGAGTCA-3', HbA primer (fw): 5'-TCACTAGCAACCT CAAACAGAC-3', HbA primer (rv): 5'-ATCCACGTTACCTTGCC-3', HbA probe: 5'-TGACTGCGGATTTTCTCAGGAGTCA-3', HbAS3 primer fw: 5'-GTGTATCCCTGGACACAAAGAT-3', HbAS3 primer (rv): 5'-GGGC TTTGACTTTGGGATTTC-3', HbAS3 probe: 5'-TTCGAAAGCTTCGGC GACCTCA-3'. Primers for HbA and HbS are identical, but probes differ by six nucleotides, and therefore it was experimentally confirmed that these two assays do not cross-react with targets. To normalize for RNA input, levels of the reference gene *RPLP0* was determined in each sample using the IDT predesigned *RPLP0* assay (Hs.PT.58.20222060). qPCR reactions were carried out on a LightCycler 480 II (Roche) using the SsoAdvanced Universal Probes Supermix (BioRad) following manufacturer's protocol and PCR conditions of 10 min at 95°C, 50 cycles of 15 s at 95°C and 60 s at 58°C. Relative mRNA levels were determined using the relative standard curve method, in which a standard curve for each gene was made from serial dilutions of the cDNA. The standard curve was used to calculate relative amounts of target mRNA in the samples relative to levels of *RPLP0*.

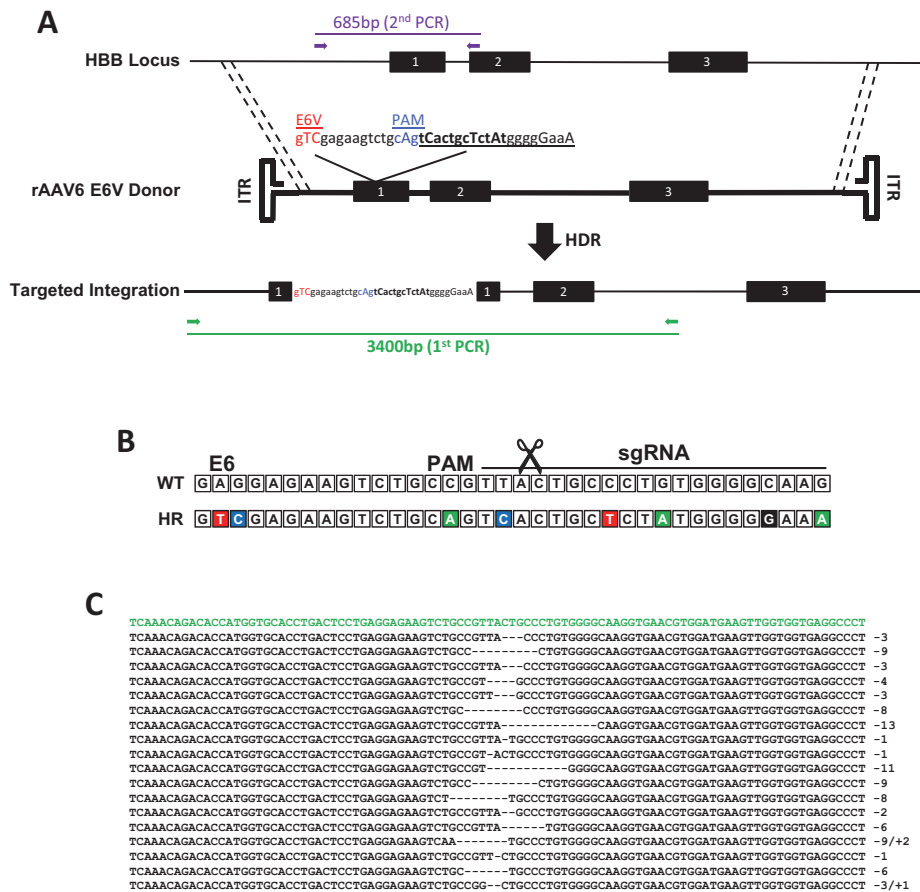
Data availability. The authors declare that the data supporting the findings of this study are available within the paper.

43. Khan, I. F., Hirata, R. K. & Russell, D. W. AAV-mediated gene targeting methods for human cells. *Nat. Protocols* **6**, 482–501 (2011).
44. Aurnhammer, C. *et al.* Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. *Hum. Gene Ther. Methods* **23**, 18–28 (2012).
45. Hendel, A. *et al.* Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Reports* **7**, 293–305 (2014).
46. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β -globin and *CCR5* genes have substantial off-target activity. *Nucleic Acids Res.* **41**, 9584–9592 (2013).
47. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
48. Boggs, D. R. The total marrow mass of the mouse: a simplified method of measurement. *Am. J. Hematol.* **16**, 277–286 (1984).



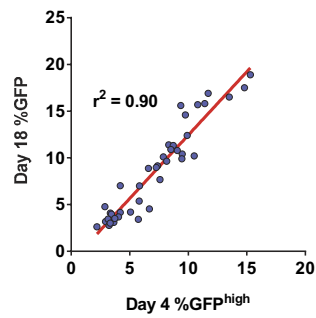
Extended Data Figure 1 | High tropism of rAAV6 for CD34⁺ HSPCs, and viability and specificity assessment of gene editing in CD34⁺ HSPCs. **a**, CD34⁺ HSPCs were transduced with a scAAV6 expressing GFP from an SFFV promoter at multiplicities of infections (MOIs) of 10,000 or 100,000 viral genomes (vg) per cell for 48 h and then analysed for percentage GFP⁺ expression by flow cytometry using a non-transduced sample to set the GFP⁺ gate at <0.1% GFP⁺ cells. scAAV was used because it eliminates second-strand synthesis as a confounder of actual transduction. Results are from two independent experiments from at least two donors and error bars represent s.d. ABM, adult bone marrow; CB, cord blood; mPB, mobilized peripheral blood. **b**, CD34⁺ HSPCs were electroporated with the *HBB* CRISPR system (mRNA or RNP delivery) or without (AAV only), and then transduced with *HBB* rAAV6 donor vectors at an MOI of 100,000 vg per cell. Day 4 after electroporation, cells were analysed by flow cytometry and live cells were gated in high forward scatter (FSC) and low side scatter (SSC). Percentage of cells in FSC/SSC gate (that is, percentage viability) is shown relative to that of mock-electroporated cells. Each data point represents a unique CD34⁺ HSPC donor. **c**, Top, sgRNA target sequences at the *HBB* on-target site and a highly complementary off-target site (Chr9:101833584–101833606) are shown. PAM sequences are underlined and red sequence highlights

the three mismatches of the off-target site. Bottom, HSPCs were electroporated with either the mRNA or RNP-based CRISPR system, and 4 days post electroporation genomic DNA was extracted and analysed for INDEL frequencies using TIDE at the on-target *HBB* and the off-target site. Results are shown as the ratio of on- to off-target activity highlighting the increased specificity of the RNP system. Averages from three different CD34⁺ HSPC donors are shown and error bars represent s.e.m. ** $P < 0.01$, unpaired Student's *t*-test. **d**, INDEL frequencies for the data presented in **c**. * $P < 0.05$, paired Student's *t*-test. **e**, Representative FACS plots showing stable GFP rates at day 18 after electroporation in donor-nuclease mismatch experiments. Mismatching nuclease and donor (red box) leads to infrequent end-capture events compared to on-target homologous recombination events observed with matched nuclease and homologous rAAV6 donor (green box). HSPCs were electroporated with 15 μ g Cas9 mRNA and either *HBB* or *IL2RG* 2'-*O*-methyl-3'-phosphorothioate-modified sgRNA, then transduced with *HBB*-GFP rAAV6 donor followed by 18 days of culture. **f**, End-capture experiments were performed in three replicate experiments each in three unique CD34⁺ HSPC donors. ns (not significant) = $P \geq 0.05$, paired Student's *t*-test. Activity of the *IL2RG* CRISPR was confirmed by quantification of INDELs at the *IL2RG* target site using TIDE analysis.



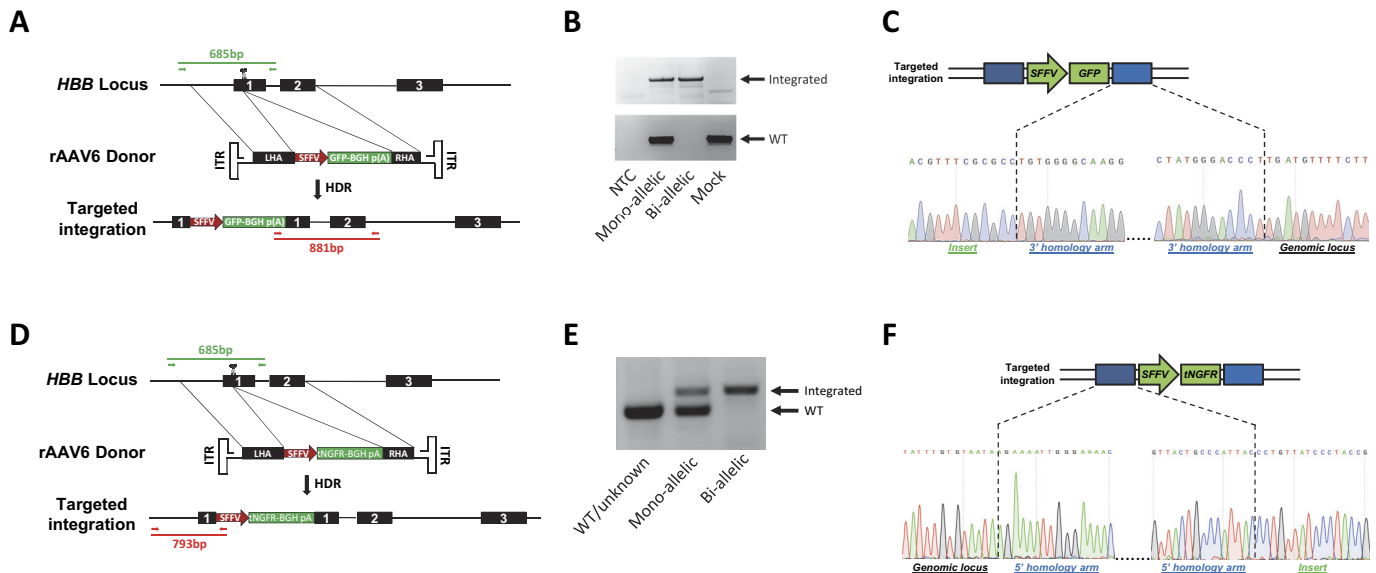
Extended Data Figure 2 | Schematic of targeting rAAV6 Glu6Val homologous donor to the *HBB* locus. **a**, The human *HBB* locus on chromosome 11 is depicted at the top of the schematic and consists of three exons (black boxes) and two introns. The rAAV6 Glu6Val donor includes the Glu to Val mutation at codon 6, which is the amino acid change causing SCD. Other SNPs (all SNPs are capitalized) were introduced to PAM site (blue) and sgRNA-binding site (bold) to prevent re-cutting following homologous recombination in HSPCs. To analyse targeted integration frequencies in HSPCs, a two-step PCR was

performed. First, a 3,400-bp in-out PCR (green) was performed followed by a nested 685-bp PCR (purple) on a gel-purified fragment from the first PCR. This second PCR fragment was cloned into TOPO vectors, which were sequenced to determine the allele genotype (wild type, INDEL or homologous recombination). **b**, The sequence of a wild-type *HBB* allele aligned with the sequence of an allele that has undergone homologous recombination. **c**, Representative INDELS from the data represented in Fig. 1d. The *HBB* reference sequence is shown in green.



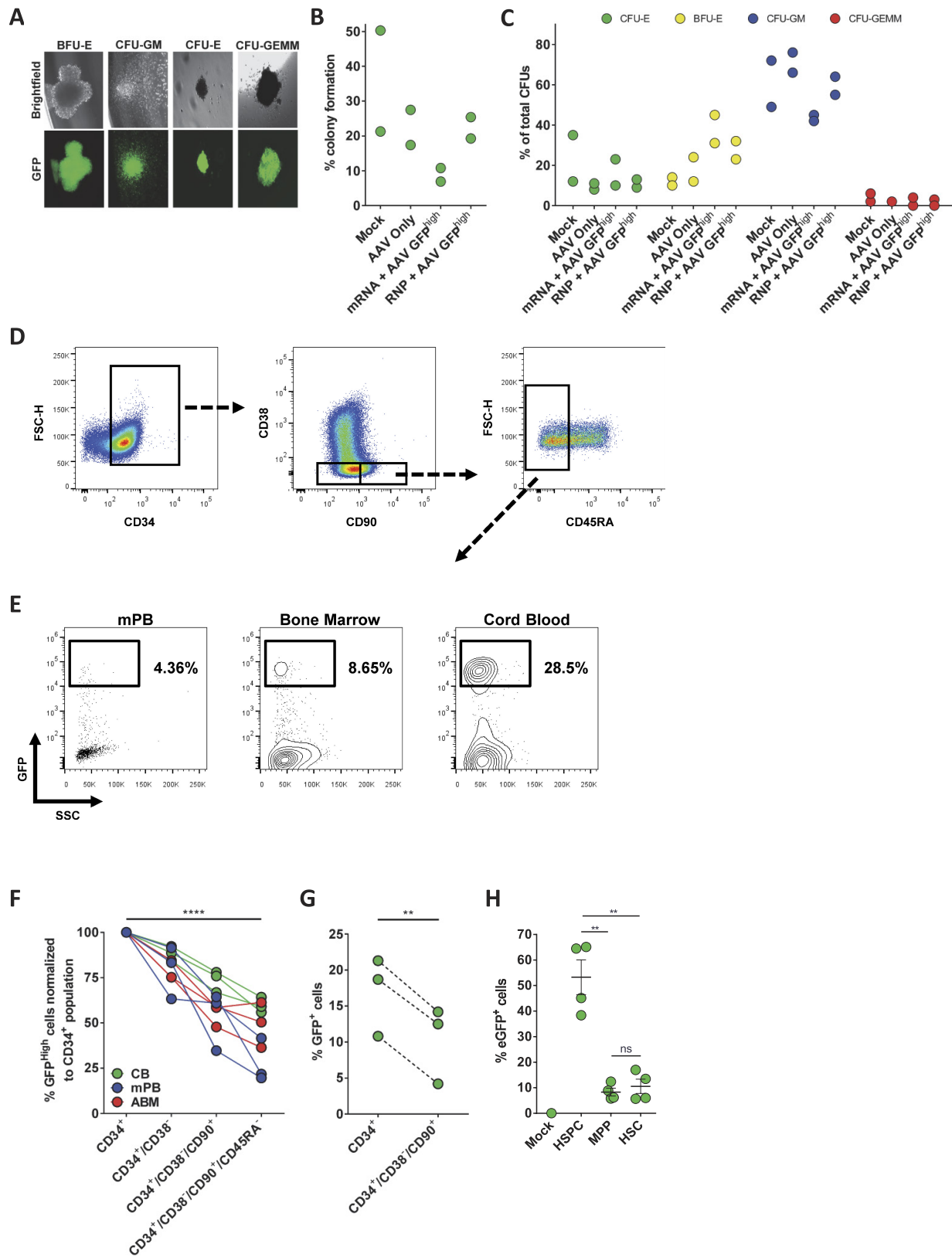
Extended Data Figure 3 | Linear regression model shows that the day 4 GFP^{high} population is a reliable predictor of targeting frequencies.

Day 4 GFP^{high} percentages (x axis) were plotted against day 18 total GFP⁺ percentages (y axis), and linear regression was performed. Data were generated from experiments including a total of 38 different CD34⁺ HSPC donors, treated with either 15 μ g or 30 μ g Cas9 RNP to generate data points with a wider distribution of targeting frequencies.



Extended Data Figure 4 | Overview of PCR genotyping of methylcellulose colonies with homologous recombination of the GFP and tNGFR donor at the *HBB* locus. **a**, The *HBB* locus was targeted by creating a DSB in exon 1 via Cas9 (scissors) and supplying a rAAV6 GFP donor template. Alleles with integrations were identified by PCR (red, 881 bp) on methylcellulose-derived colonies using an in-out primer set. Wild-type alleles were identified by PCR (green, 685 bp) using primers flanking the sgRNA target site. **b**, Representative genotyping PCRs showing mono- and biallelic clones as well as a clone derived from mock-treated cells. NTC, non-template control (see Supplementary Fig. 1a for uncropped gel). **c**, Representative Sanger sequence chromatograms for junctions between right homology arm (blue) and

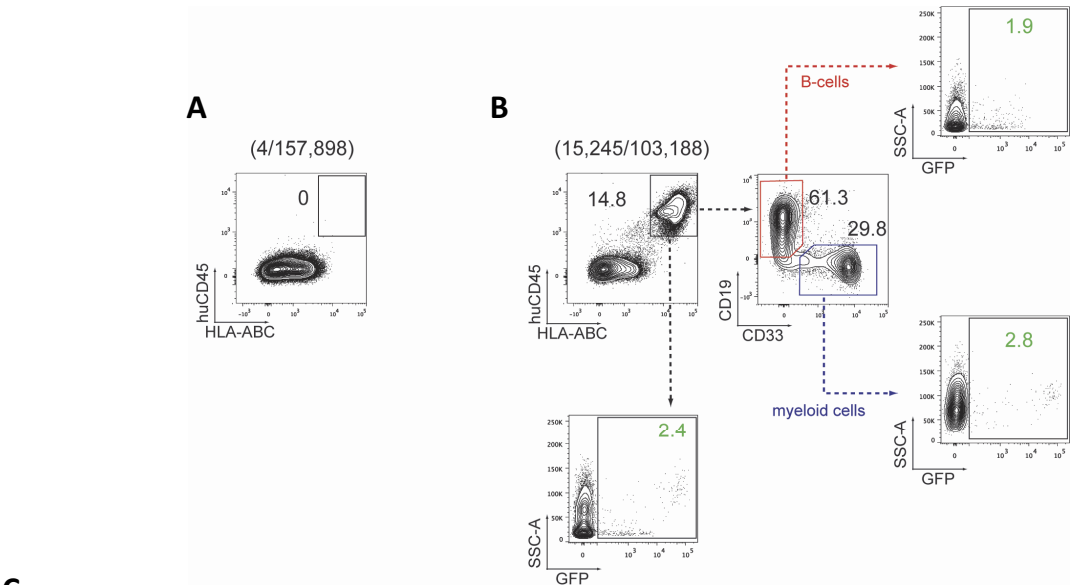
insert (green) or genomic locus, highlighting seamless homologous recombination. **d**, The *HBB* locus was targeted by creating a DSB in exon 1 via Cas9 (scissors) and supplying a rAAV6 tNGFR donor template. Genotypes were assessed by a three-primer genotyping PCR on methylcellulose-derived colonies using an in-out primer set (red, 793 bp) and a primer set flanking the sgRNA target site (green, 685 bp). Note that the two forward primers are the same. **e**, Representative genotyping PCRs showing a wild-type/unknown, mono-, and biallelic clone (see Supplementary Fig. 1b for uncropped gel). **f**, Representative Sanger sequence chromatograms for junctions between left homology arm (in blue) and insert (in green) or genomic locus highlighting seamless homologous recombination.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Haematopoietic progenitor CFU assay and targeting in different HSPC subpopulations. **a**, GFP^{high} HSPCs were single-cell-sorted into 96-well plates containing methylcellulose. Representative images from fluorescence microscopy show lineage-restricted progenitors (BFU-E, CFU-E, CFU-GM) and multipotent progenitors (CFU-GEMM) with GFP expression. **b**, CFUs were counted 14 days after sorting and shown relative to the total number of cells sorted (percentage colony formation) ($n = 2$ different HSPC donors). **c**, Colonies were scored according to their morphology: (1) CFU-erythroid (CFU-E); (2) burst forming unit-erythroid (BFU-E); (3) CFU-granulocyte/macrophage (CFU-GM); and (4) CFU-granulocyte/erythrocyte/macrophage/megakaryocyte (CFU-GEMM) ($n = 2$ different HSPC donors). **d**, Representative FACS plots at day 4 after electroporation of CD34⁺ HSPCs showing the gating scheme for analysing targeting frequencies in different HSPC subsets (Extended Data Fig. 5f). Cells were immunophenotyped for CD34, CD38, CD90 and CD45RA expression and relevant FACS gates are indicated. **e**, Representative FACS plots showing GFP^{high} cells in the CD34⁺ CD38⁻ CD90⁺ CD45RA⁻ population of HSPCs derived from mobilized peripheral blood, bone marrow, or cord blood. **f**, 500,000 HSPCs isolated from mobilized peripheral blood, adult bone marrow, or cord blood were electroporated with RNP and transduced with GFP rAAV6 donor. At day 4 after electroporation, cells

were phenotyped by flow cytometry for the cell surface markers CD34, CD38, CD90 and CD45RA (Extended Data Fig. 5d, e). Percentage GFP^{high} cells in the indicated subpopulations are shown (data points represent unique donors, $n = 3$ per HSPC source), **** $P < 0.0001$, paired Student's t -test. **g**, CD34⁺ or CD34⁺ CD38⁻ CD90⁺ cells were sorted directly from freshly isolated cord blood CD34⁺ HSPCs, cultured overnight, and then electroporated with RNP and transduced with GFP rAAV6. Bars show average percentage GFP⁺ cells at day 18 after electroporation. ($n = 3$ from different HSPC donors), ** $P < 0.01$, paired Student's t -test. **h**, Multipotent progenitor (MPP) (CD34⁺ CD38⁻ CD90⁻ CD45RA⁻) and HSC (CD34⁺ CD38⁻ CD90⁺ CD45RA⁻) populations were sorted from fresh cord-blood-derived CD34⁺ HSPCs and immediately after sorting, cells were transduced with scAAV6-SFFV-GFP at an MOI of 100,000 vg per cell along the bulk HSPC population. scAAV6 was used because it eliminates second-strand synthesis as a confounder of actual transduction, although the activity of the SFFV promoter may not be equivalent in each population, thus potentially underestimating the degree of transduction of MPPs and HSCs. Two days later, transduction efficiencies were measured by FACS analysis of GFP expression using non-transduced cells (mock) to set the GFP⁺ gate. Error bars represent s.e.m., $n = 4$, two different HSPC donors. ** $P < 0.01$; NS, not significant = $P \geq 0.05$, unpaired t -test with Welch's correction.

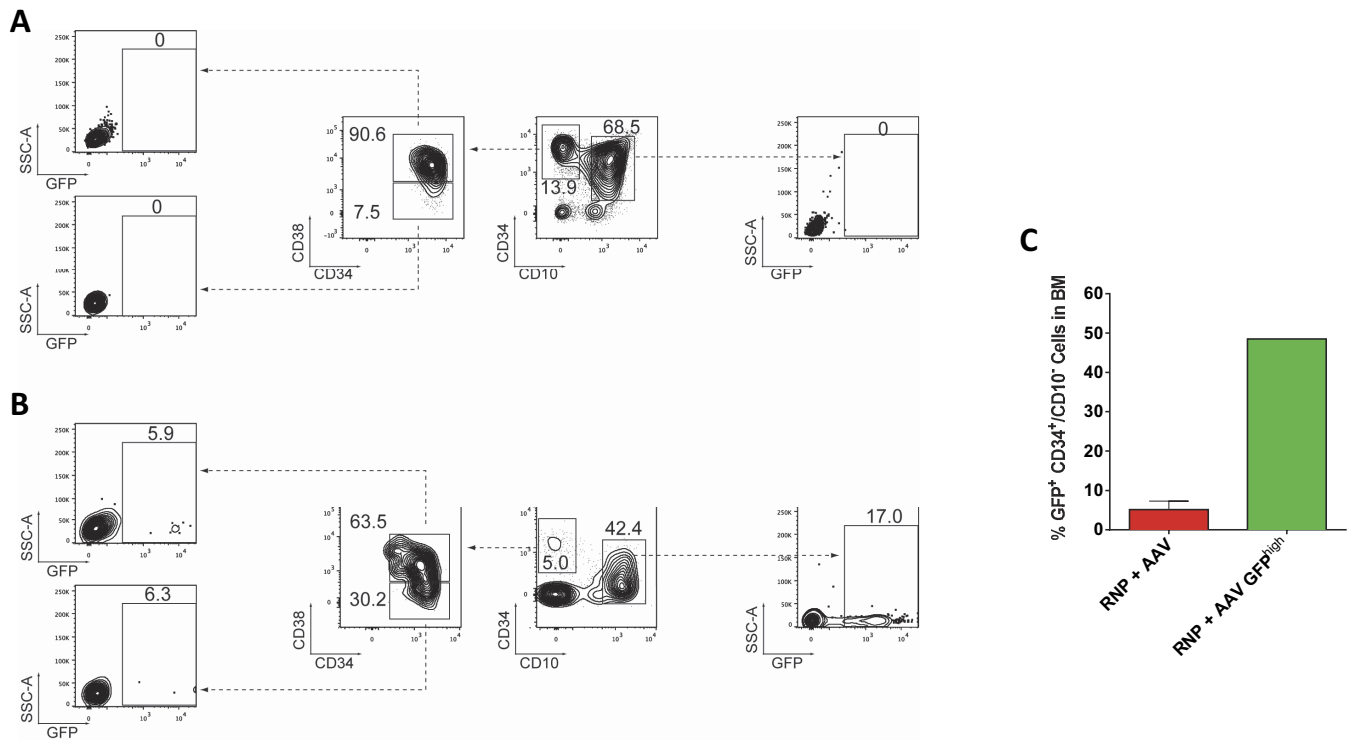


Sample Name	% engraftment (average per mouse)	% GFP ⁺ cells among human cells (average per mouse)	Total live cells injected per mouse (average per mouse)	Average total HSCs injected per mouse (CD34 ⁺ /CD38 ⁻ /CD90 ⁺ /CD45RA ⁻)	Estimated total modified human cells in BM per mouse at Week 16
RNP + AAV	6.99 (+/-1.8)	3.46 (+/-0.82)	500,000	24,000 (+/-1100)	270,000 (+/-33,000)
RNP + AAV GFP ^{high}	1.67 (+/-1.0)	78.8 (+/-10.6)	350,000 100,000 x1 250,000 x2 500,000 x3	2,900 (+/-770)	1,460,000 (+/-390,000)

Extended Data Figure 6 | Analysis of human engraftment.

a, Representative FACS plot from the analysis of the bone marrow of a control mouse not transplanted with human cells. Mice were euthanized and bone marrow was collected from femur, tibia, hips, humerus, sternum and vertebrae. Cells were subject to Ficoll density gradient to isolate mononuclear cells, which were analysed for human engraftment by flow cytometry. Human engraftment was delineated as hCD45/HLA-ABC double positive. From a total of 157,898 cells, 4 were found within the human cell gate of a non-injected control mouse, showing the very limited background human staining. **b**, Representative FACS plots showing gating scheme for analyses of NSG mice transplanted with human cells and analysed as described in **a**. Representative plots are from one mouse from the RNP plus rAAV6 experimental group. As above, human engraftment was delineated as hCD45/HLA-ABC double positive. B cells were marked by CD19 expression, and myeloid cells identified by CD33 expression. GFP expression was analysed in total human cells (2.4%), B cells (1.9%) and myeloid cells (2.8%). The GFP brightness in B cells is lower than in

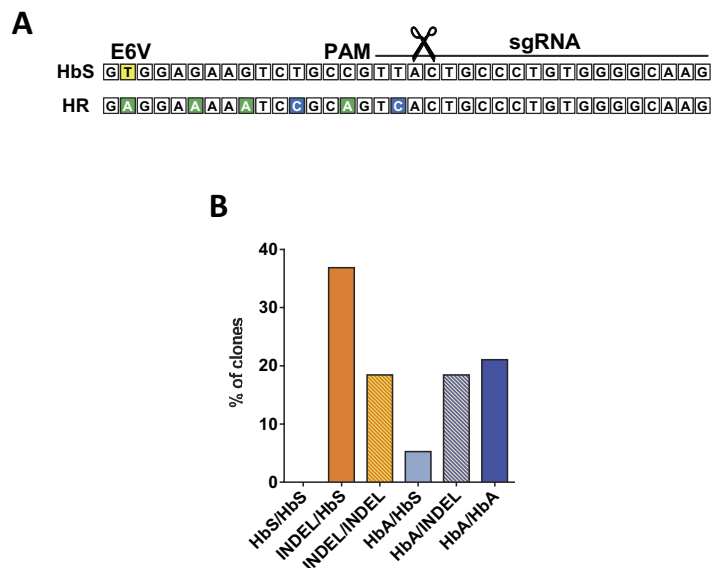
myeloid cells, suggesting that the SFFV promoter is not as active in the B-cell lineage compared to the myeloid lineage (see also Fig. 3a). **c**, Overview of engraftment for RNP plus AAV and RNP plus AAV GFP^{high} experimental groups. Average engraftment frequencies and percentage GFP⁺ human cells \pm s.e.m. are shown. Total number of cells transplanted was the same (500,000) for all mice in the RNP group, whereas in the GFP^{high} group, one mouse was transplanted with 100,000 cells, two mice with 250,000 cells, and three mice with 500,000 cells. The total number of HSCs transplanted per mouse (\pm s.e.m.) was calculated based on the frequencies of GFP⁺ cells in the CD34⁺ CD38⁻ CD90⁺ CD45RA⁻ subset analysed by flow cytometry (see Extended Data Fig. 5f) directly before injection. The total number of modified human cells in the bone marrow at week 16 after transplant per mouse (\pm s.e.m.) was estimated based on calculations presented in the Methods. This shows that the enrichment not only resulted in a higher percentage of edited cells (column 3) but also resulted in an absolute higher number (column 6) of edited cells.



Extended Data Figure 7 | Genome-edited human HSPCs in the bone marrow of NSG mice at week 16 after transplantation.

a, b, Representative FACS plots from the analysis of NSG mice from the mock (**a**) or RNP plus AAV (**b**) experimental group at week 16 after transplantation. Mice were euthanized and bone marrow was obtained, MNCs were isolated via Ficoll density gradient, after which human CD34⁺ cells were enriched by magnetic-activated cell sorting (MACS), and finally cells were stained with anti-CD34, anti-CD38 and anti-CD10

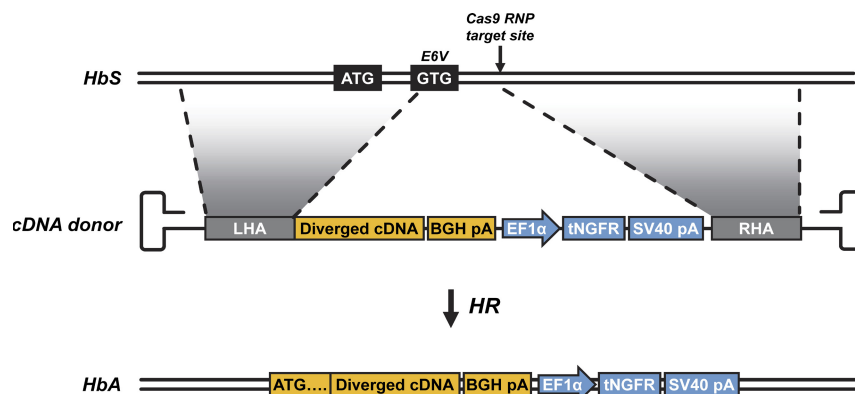
antibodies to identify human GFP⁺ cells in the CD34⁺ CD10⁻ and CD34⁺ CD10⁻ CD38⁻ populations (note that CD10 was included as a negative discriminator for immature B cells). **c,** Collective data from the analysis of GFP⁺ cells in the human CD34⁺ CD10⁻ population from the RNP plus AAV ($n = 11$) and RNP plus AAV GFP^{high} ($n = 6$) experimental groups. For the RNP plus AAV GFP^{high} group, cells from all six mice were pooled before analysis and thus, no error bar is available. Error bar on RNP group represents s.e.m.



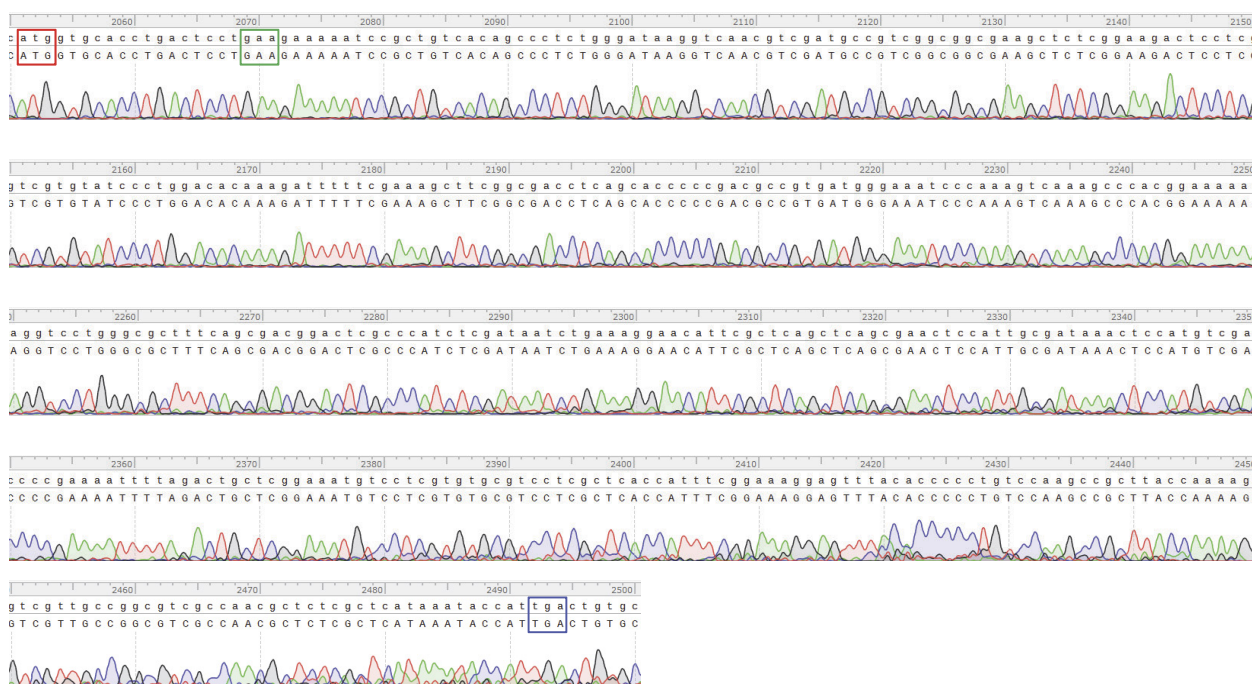
Extended Data Figure 8 | Correction of the sickle cell mutation in patient-derived CD34⁺ HSPCs. a, Schematic overview of the sequence of the sickle cell allele aligned with the sequence of an allele that has undergone homologous recombination using the corrective SNP donor. The Glu6Val mutation in patients with SCD (A>T) is highlighted in yellow. The sgRNA recognition sequence, the PAM site and the cut site (scissors) are shown. The donor carries synonymous nucleotide changes between the sickle nucleotide and the cut site to avoid premature crossover

during homologous recombination. Synonymous changes are also added to the PAM and an early nucleotide in the sgRNA target site to avoid subsequent re-cutting and potential inactivation of the corrected allele. **b,** HSPCs from two different patients with SCD were targeted with the corrective SNP donor and seeded in methylcellulose. After 14 days, in-out PCR amplicons from a total of 38 clones were sequenced and genotypes were extracted from sequence chromatograms.

A

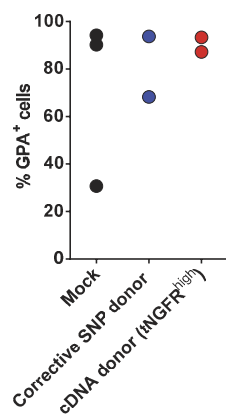


B



Extended Data Figure 9 | Targeting *HBB* with a cDNA donor and a tNGFR expression cassette. **a**, Schematic representation of the AAV6 donor encoding a functional *HBB* anti-sickling cDNA (Gly16Asp, Glu22Ala, Thr87Gln) followed by an expression cassette for tNGFR. The left homology arm stops just before the sickle mutation (A→T) followed by the remaining *HBB* cDNA, which has been diverged from the endogenous sequence by introducing synonymous mutations at codon wobble positions. The *HBB* cDNA expression cassette is followed by an EF1 α promoter driving tNGFR expression (*HBB* cDNA-EF1 α -tNGFR),

which constitutes a clinically compatible expression cassette enabling enrichment and tracking of modified cells. **b**, Chromatogram from sequencing of an in-out PCR amplicon on CD34⁺ cells derived from patients with SCD, electroporated with *HBB* Cas9 RNP and transduced with rAAV6 *HBB* cDNA donor. PCR was performed on genomic DNA extracted 4 days after electroporation of a bulk sample. Chromatogram shows the sequence of the full *HBB* cDNA with start codon, sickle-cell codon position (containing a corrected Glu codon), and stop codon highlighted in red, green and blue boxes, respectively.



Extended Data Figure 10 | Edited HSPCs from patients with SCD differentiate into erythrocytes that express glycophorin A. CD34⁺ HSPCs derived from patients with SCD were edited with *HBB* Cas9 RNP and either the corrective SNP donor or the cDNA donor. Four days after electroporation, cells edited with the cDNA donor were sorted for tNGFR⁺ cells. This population as well as the populations edited with the corrective SNP donor and mock-electroporated cells were subjected to a 21-day erythrocyte differentiation protocol, followed by staining for glycophorin A (GPA). All data points within experimental groups are derived from experiments in cells from different patients with SCD, $n = 3$ (mock) and $n = 2$ (SNP and cDNA donor).

LKB1 loss links serine metabolism to DNA methylation and tumorigenesis

Filippos Kottakis^{1,2,3}, Brandon N. Nicolay^{1,3}, Ahlima Roumane^{1,2,3}, Rahul Karnik^{4,5,6}, Hongcang Gu^{4,5,6}, Julia M. Nagle^{1,2,3}, Myriam Boukhali^{1,3}, Michele C. Hayward⁷, Yvonne Y. Li^{8,9}, Ting Chen^{8,9,10}, Marc Liesa^{11,12}, Peter S. Hammerman^{8,9,13}, Kwok Kin Wong^{8,9,10}, D. Neil Hayes⁷, Orian S. Shirihai^{11,12}, Nicholas J. Dyson^{1,3}, Wilhelm Haas^{1,3}, Alexander Meissner^{4,5,6} & Nabeel Bardeesy^{1,2,3}

Intermediary metabolism generates substrates for chromatin modification, enabling the potential coupling of metabolic and epigenetic states. Here we identify a network linking metabolic and epigenetic alterations that is central to oncogenic transformation downstream of the liver kinase B1 (LKB1, also known as STK11) tumour suppressor, an integrator of nutrient availability, metabolism and growth. By developing genetically engineered mouse models and primary pancreatic epithelial cells, and employing transcriptional, proteomics, and metabolic analyses, we find that oncogenic cooperation between LKB1 loss and KRAS activation is fuelled by pronounced mTOR-dependent induction of the serine–glycine–one–carbon pathway coupled to S-adenosylmethionine generation. At the same time, DNA methyltransferases are upregulated, leading to elevation in DNA methylation with particular enrichment at retrotransposon elements associated with their transcriptional silencing. Correspondingly, LKB1 deficiency sensitizes cells and tumours to inhibition of serine biosynthesis and DNA methylation. Thus, we define a hypermetabolic state that incites changes in the epigenetic landscape to support tumorigenic growth of LKB1-mutant cells, while resulting in potential therapeutic vulnerabilities.

Substrates and inhibitors of chromatin-modifying enzymes are generated in intermediary metabolism, so changes in nutrient availability and utilization can influence epigenetic regulation^{1,2}. Importantly, recent studies have indicated that the interplay between metabolism and epigenetics can serve as a programmed switch in cell states. For example, mouse embryonic stem cell differentiation is promoted by succinate-mediated inhibition of histone demethylases (HDMs) and TET DNA demethylases³, or by decreased S-adenosyl-methionine (SAM) levels leading to loss of histone H3K4 methylation⁴. Moreover, aberrant metabolic activity can produce pathological effects by altering chromatin regulation. Most notably, mutations in the genes encoding the isocitrate dehydrogenase (IDH1 and IDH2) enzymes lead to the generation of 2-hydroxyglutarate, which inhibits HDMs and TETs and thereby alters DNA and histone methylation—changes that have been implicated in overriding cell differentiation and promoting tumorigenesis⁵. Whether this paradigm extends more generally to other oncogenic mutations remains unclear, and this question has implications for understanding cancer pathogenesis and developing improved treatments. Here, we demonstrate that dynamic exchange between metabolism and chromatin regulation contributes to pancreatic tumorigenesis driven by mutation of the LKB1 serine–threonine kinase.

LKB1 is mutationally inactivated in a range of sporadic cancers, including pancreatic carcinomas^{6–8}. Additionally, germline mutations in LKB1 cause Peutz-Jeghers syndrome, which comprises gastrointestinal polyps and a high incidence of gastrointestinal tract carcinomas (for example, an approximately 100-fold increase in pancreatic cancer)^{9,10}. Cancers with LKB1 mutations tend to exhibit aggressive clinical features and different therapeutic sensitivity from cancers without these

mutations^{11–14}. LKB1 directly activates a family of 14 kinases related to AMP-activated protein kinase (AMPK), many of which are coupled to nutrient sensing and broadly reprogram cell metabolism¹⁵. Thus, metabolic rewiring is thought to be a driver of tumorigenesis after LKB1 loss. We now identify an LKB1-regulated program that links metabolic alterations to control of the epigenome and is involved in malignant growth. Our results provide evidence that coupled metabolic and epigenetic states have a more general role in cancer pathogenesis and suggest therapeutic strategies that could target these intersecting processes.

Synergy between LKB1 and KRAS mutations

LKB1 inactivation frequently coincides with mutations in the RAS–RAF pathway in human cancers and these genetic alterations cooperate to drive tumorigenesis in genetically engineered mouse models (GEMMs)^{6,11,14,16}. We examined the interactions between oncogenic KRAS^{G12D} and deletion of LKB1 in adult pancreatic ducts using a tamoxifen-inducible GEMM (Extended Data Fig. 1a). The combined alterations resulted in pancreatic cancers by 20–25 weeks, whereas the individual mutations had no pathological effects at this age (Fig. 1a and Extended Data Fig. 1b). To investigate the mechanisms of tumorigenesis, we isolated primary pancreatic ductal epithelial cells from mice with conditional KRAS^{G12D} and LKB1 alleles ($n = 2$ lines per genotype) and transduced them with adenoviruses expressing Cre and/or Flp recombinase to generate KRAS^{G12D/+}, LKB1^{-/-} and KRAS^{G12D/+};LKB1^{-/-} cells (K, L and KL cells, respectively) as well as wild-type parental lines (Extended Data Fig. 1c). Only KL cells were tumorigenic following injection into severe combined immunodeficient (SCID) mice or growth in soft agar, and tumorigenicity was blocked by restoration of

¹Cancer Center, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. ²Center for Regenerative Medicine, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. ³Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA. ⁴Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁵Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. ⁶Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁷UNC, Lineberger Comprehensive Cancer Center, Chapel Hill, North Carolina 27599, USA. ⁸Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. ⁹Department of Medical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts 02215, USA. ¹⁰Belfer Institute for Applied Cancer Science, Dana Farber Cancer Institute, Boston, Massachusetts 02215, USA. ¹¹Evans Center for Interdisciplinary Research, Department of Medicine, Mitochondria ARC, Boston University School of Medicine, Boston, Massachusetts 02118, USA. ¹²Department of Medicine, Division of Endocrinology, Diabetes and Hypertension, UCLA David Geffen School of Medicine, Los Angeles, California 90095, USA. ¹³Cancer Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

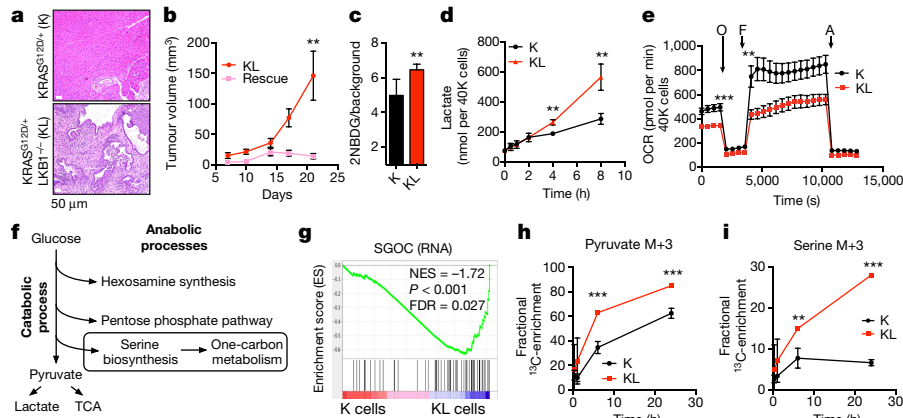


Figure 1 | LKB1 inactivation synergizes with KRAS^{G12D} to potentiate glycolysis, serine metabolism, and tumorigenesis. **a**, Representative pancreas histology of the indicated genotypes of mice at 20–25 weeks ($n = 4$ per genotype). **b**, Subcutaneous tumour growth of KL cells expressing empty vector (KL) or LKB1 (rescue) ($n = 8$ per group). **c**, **d**, Ductal cells tested for glucose uptake (**c**; $n = 6$, independent replicates), and lactate secretion (**d**; $n = 3$). **e**, Oxygen consumption rates

in K and KL cells under nutrient-replete conditions ($n = 21$). **f**, Fates of glycolytic intermediates. **g**, GSEA showing enrichment of serine–glycine–one-carbon network¹⁸ (K cells, $n = 3$; KL cells, $n = 4$). NES, normalized enrichment score. **h**, **i**, Isotopomer abundance of U[¹³C]glucose-derived M+3 pyruvate (**h**) or serine (**i**; $n = 3$, biological replicates). Data pooled from three (**e**) or representative of two (**d**) experiments. Error bars, s.e.m. (**b**), s.d. (**c**, **d**, **g**, **h**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

wild-type LKB1 (Fig. 1b and Extended Data Fig. 1d–g). *In vitro*, KL cells showed greater proliferation than K cells, and both showed greater proliferation than wild-type or L cells (Extended Data Fig. 1h–j). Thus, primary ductal cells provide a tractable *in vitro* system to study mechanisms of epithelial cell transformation arising from LKB1 inactivation.

Focusing on the metabolic alterations provoked by loss of LKB1, we found that KL cells exhibited an approximately 30% increase in glucose uptake compared to K cells, and showed marked elevations in levels of the GLUT1 transporter and ATP (Fig. 1c and Extended Data Fig. 1k, l). Lactate levels were elevated in KL cells, whereas oxygen consumption and citrate levels were reduced (Fig. 1d, e and Extended Data Fig. 1m). Moreover, KL cells showed heightened sensitivity to acute glucose deprivation and to inhibition of glycolysis using the glucose analogue 2-deoxyglucose, the pyruvate dehydrogenase kinase inhibitor dichloroacetate, or the lactate dehydrogenase inhibitor galloflavin (Extended Data Fig. 1n–q). Importantly, neither KRAS^{G12D} nor LKB1 inactivation alone promoted significant alterations in glucose metabolism

(Extended Data Fig. 1r–t). Thus, these genetic lesions acted synergistically to potentiate glycolysis while rendering cells highly dependent on glucose availability.

These data suggested that an increased supply of glycolytic intermediates was available for anabolic processes to support the growth of KL cells (Fig. 1f). Notably, gene set enrichment analysis (GSEA) of RNA-sequencing (RNA-seq) and quantitative proteomics¹⁷ data indicated that KL cells are enriched for glycolytic enzymes and for networks that connect glycolytic intermediates to one-carbon metabolism, with serine–glycine–threonine and folate metabolism scoring highly among the induced pathways (Supplementary Data Table 1). There was particularly striking enrichment of a 64-gene signature defining the entire serine–glycine–one carbon (SGOC) network¹⁸, indicating strong coordinate activation of these pathways (Fig. 1g and Extended Data Fig. 2a). Accordingly, the use of uniformly carbon-13-labelled glucose (U[¹³C]glucose) demonstrated that KL cells show augmented production of glucose-derived pyruvate and lactate and an even more

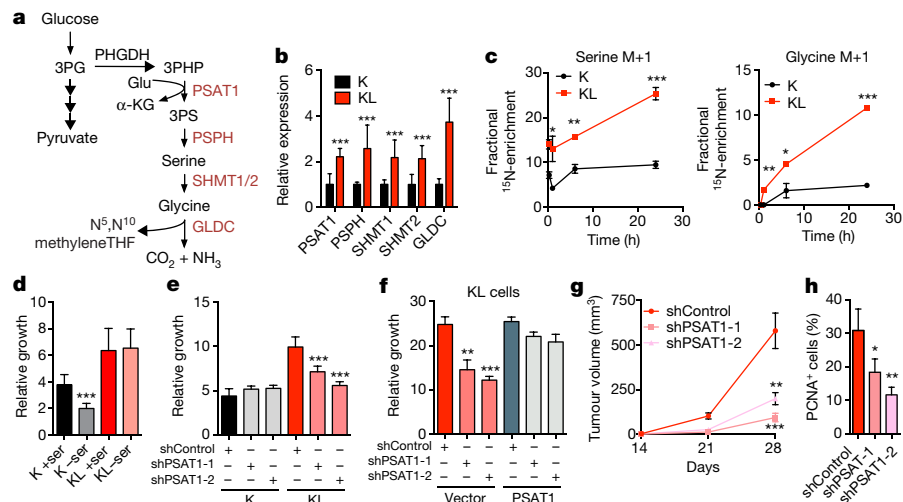


Figure 2 | Activation of *de novo* serine biosynthesis supports growth of LKB1-deficient cells. **a**, Serine biosynthesis pathway. Red, upregulated in KL cells. **b**, Serine pathway gene expression ($n = 8$ per genotype). **c**, Isotopomer abundance of [¹⁵N]glutamine-derived M+1 serine and glycine ($n = 3$, biological replicates). **d**, **e**, Three-day growth of ductal cells cultured with or without 0.4 mM serine (**d**; $n = 20$) or transduced with the indicated shRNAs (**e**; $n = 6$). **f**, Six-day proliferation of KL cells

transduced with the indicated shRNAs and expression constructs ($n = 3$). **g**, Subcutaneous growth of tumours from KL cells transduced with the indicated shRNAs ($n = 12$ tumours per group). **h**, Proportion of CK19⁺ tumour cells that are PCNA⁺ (shControl $n = 4$, shPSAT1-1 $n = 4$, shPSAT1-2 $n = 3$, representative tumours). Data pooled from four (**b**) or representative of two (**e**, **f**) or four (**d**) experiments. Error bars: s.d. (**b**–**f**), s.e.m. (**g**, **h**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

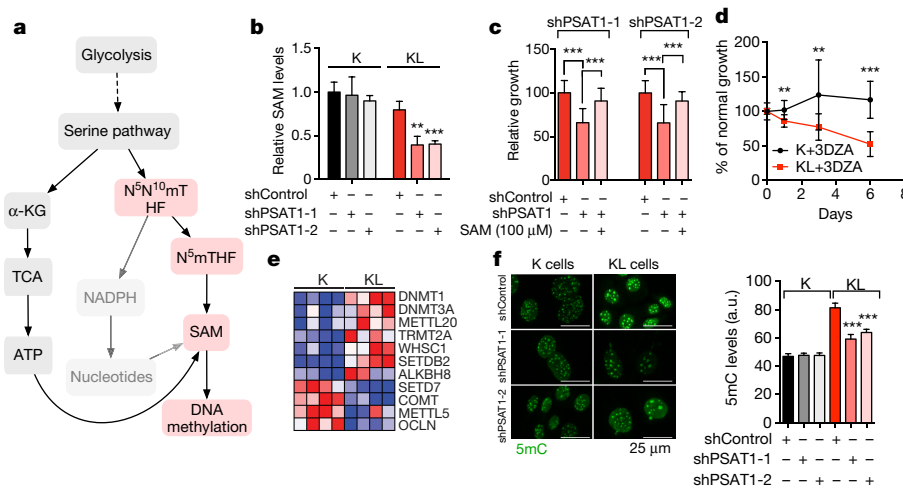


Figure 3 | Activation of the SGOC network promotes DNA methylation in LKB1-deficient cells. a, SGOC network. **b**, SAM levels in ductal cells transduced with the indicated shRNAs ($n = 4$). **c**, Three-day growth of KL cells with or without SAM supplementation or PSAT1 knockdown ($n = 16$). **d**, Proliferation of ductal cells treated with 3-deazaadenosine

(3-DZA) ($n = 12$). **e**, Heatmap of differentially regulated SAM-using enzymes (proteomics). **f**, 5mC in ductal cells (77–177 cells). Data pooled from two (**c**, **d**, **f**) or representative of two (**b**) experiments. Error bars: s.d. (**b–d**), s.e.m. (**f**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

pronounced increase in serine and glycine biosynthesis rates without changes in total levels of these amino acids (Fig. 1h, i and Extended Data Fig. 2b–d).

Serine pathway dependence of KL cells

Multiple serine pathway enzymes (PSAT1, PSPH, SHMT1 and SHMT2) were upregulated in KL cells, whereas restoration of LKB1 expression reversed these changes, broadly suppressed the entire SGOC network, and reduced serine biosynthesis (Fig. 2a, b and Extended Data Fig. 2e–g). PSAT1 catalyses the transamination of 3-phosphohydroxy-pyruvate (3PHP) to 3-phosphoserine (3PS), with glutamate as the nitrogen donor and α -ketoglutarate (α -KG) as a secondary product (Fig. 2a). Consistent with elevated PSAT1 activity, [15 N]glutamine labelling revealed a marked increase in nitrogen incorporation into serine and glycine in KL cells (Fig. 2c). Thus, LKB1 restricts serine metabolism.

Consistent with this finding, KL cells were unaffected by culturing in serine-free medium, whereas K cells showed a roughly 40% decrease in proliferation, as did KL cells in which LKB1 was expressed (Fig. 2d and Extended Data Fig. 3a). On the other hand, KL cells were specifically sensitive to PSAT1 knockdown, exhibiting reduced proliferation in normal medium, restored dependency on exogenous serine, and impaired colony formation in soft agar (Fig. 2e and Extended Data Fig. 3b–d). Introduction of short hairpin (sh)RNA-resistant human PSAT1 cDNA rescued these phenotypes (Fig. 2f and Extended Data Fig. 3e). Moreover, PSAT1 knockdown strongly inhibited subcutaneous tumour growth by KL cells (Fig. 2g, h and Extended Data Fig. 3f, g). Notably, PSAT1 knockdown had no effect on the proliferation or tumorigenicity of cell lines from pancreatic cancer GEMMs with wild-type LKB1 (KRAS^{G12D}-p53 null, KPC and KRAS^{G12D}-CDKN2A null, KIC)¹⁹, supporting the idea that these effects result from LKB1 loss rather than being secondary to high proliferation rates (Extended Data Fig. 3b, h–k). Thus, increased serine biosynthesis pathway activity is required to drive oncogenic transformation in KL cells.

Serine biosynthesis fuels DNA methylation

The serine biosynthesis pathway can fuel a range of anabolic processes, including generation of N^5,N^{10} -methylene-tetrahydrofolate (THF), which supports NADPH production, redox homeostasis, and nucleotide biosynthesis²⁰. PSAT1 knockdown failed to increase levels of reactive oxygen species (ROS) in KL cells and, correspondingly, proliferation was not restored by the ROS scavenger *N*-acetylcysteine (NAC) or by nucleoside supplementation (Extended Data Fig. 4a–d).

Similar results were obtained after treatment with aminooxyacetate (AOA), an inhibitor of PSAT1 and other aminotransferases (Extended Data Fig. 4e, f). Thus, the requirement of KL cells for PSAT1 appears to be independent of its roles in redox homeostasis or maintenance of nucleoside pools.

Serine metabolism can also fuel the methionine salvage pathway (Fig. 3a), which is the main mechanism of production for the methyl donor *S*-adenosyl methionine (SAM)^{21,22}. Notably, multiple enzymes that channel serine metabolism intermediates into the methionine salvage pathway and that contribute to SAM biosynthesis were upregulated in KL cells (Extended Fig. 4g). Accordingly, PSAT1 knockdown reduced SAM levels in KL cells but not in K cells (Fig. 3b). Furthermore, SAM supplementation mitigated the proliferation defects caused by PSAT1 ablation in KL cells (Fig. 3c). Conversely, suppression of SAM biosynthesis by inhibition of *S*-adenosyl-homocysteine-hydrolase (SAH) with 3-deazaadenosine (3DZA), or of methionine-adenosyl-transferase (MAT) with cycloleucine, slowed the growth of KL cells, whereas K cells remained unaffected (Fig. 3d and Extended Data Fig. 4h). Thus, activation of *de novo* serine biosynthesis contributes to SAM production to support growth of LKB1-deficient cells, possibly by providing active methyl groups in the form of N^5 -methyl-THF, or by promoting ATP generation (for example, via PSAT1-dependent TCA cycle anaplerosis)^{22,23}.

SAM is the substrate for methylation of lipids, DNA, RNA, metabolites and proteins. Examination of our RNA-seq and proteomics data sets for the expression of the 183 annotated SAM-dependent methyltransferase enzymes revealed significant upregulation of the DNA methyltransferases DNMT1 and DNMT3A in KL cells, whereas few of the other enzymes showed any changes (Fig. 3e, Extended Data Fig. 5a and Supplementary Data Table 2). Quantitative PCR and immunoblot analyses verified the regulation of expression of these enzymes by LKB1 (Extended Data Fig. 5b, c). Importantly, inhibition of DNA methylation using 5-aza-2-deoxycytidine (decitabine) or the non-nucleoside DNMT inhibitor RG108 promoted SAM accumulation in KL cells but not in K cells (Extended Data Fig. 5d), suggesting that DNMTs are major SAM-utilizing enzymes in the context of LKB1 deficiency.

Accordingly, immunofluorescence and DNA dot blot analysis demonstrated that KL cells had a significant increase in 5-methylcytosine (5mC) as compared to K cells, an effect that was reversed by LKB1 rescue (Extended Data Fig. 5e–g). By contrast, there were no consistent changes in global levels of a series of histone methylation marks, in 5-hydroxy-methylcytosine, or in N_6 -methyladenosine (Extended Data Fig. 5h–i and data not shown). Notably, PSAT1 knockdown suppressed

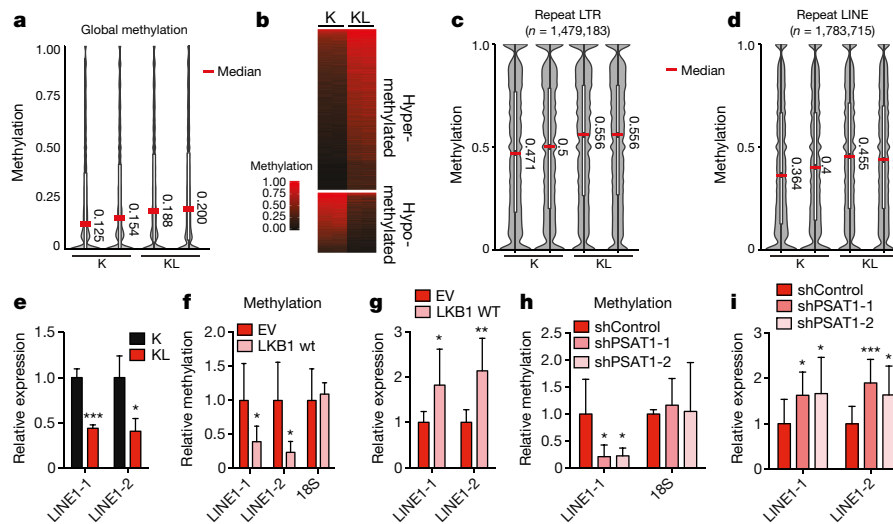


Figure 4 | The LKB1–PSAT1 pathway controls methylation and expression of retrotransposons. **a**, Distribution of methylation levels in ductal cells. Median values are indicated. **b**, Heatmap of differentially methylated tiles. **c**, **d**, Distribution of methylation levels in LINE1 (**c**) and LTR (**d**) repeats. **e**, Expression of LINE1 in ductal cells, tested using two independent primer sets ($n = 3$). **f**, **g**, KL cells transduced with EV or LKB1

assessed for methylation of LINE1 and 18S (**f**; $n = 6$, LINE1-1 and 18S; $n = 4$, LINE1-2), and LINE1 expression (**g**; $n = 6$). **h**, **i**, KL cells transduced with the indicated shRNAs were assessed for LINE1 and 18S methylation (**h**; $n = 6$) and LINE1 expression (**i**; $n = 10$). Data pooled from two (**f**, **h**) or representative of two (**e**, **g**, **i**) experiments. Error bars: s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

DNA methylation in KL cells, and this effect was reversed by shRNA-resistant PSAT1 cDNA, while K and KPC controls were unaffected (Fig. 3f and Extended Data Fig. 6a–c). Comparison of all four genotypes using our set of isogenic ductal cells confirmed that serine pathway activity and DNA methylation were specifically potentiated upon dual KRAS^{G12D} expression and LKB1 loss (Extended Data Fig. 6d–g).

The expression of a kinase-dead (KD) LKB1 mutant confirmed that the activation of key transcriptional and metabolic circuits in KL cells was due to loss of LKB1 kinase activity (Extended Data Fig. 7a–e). Accordingly, KL cells showed reduced activity of AMPK, an important target of the LKB1 kinase, and consequent activation of mechanistic target of rapamycin complex I (mTORC1) (Extended Data Fig. 7f, w). Moreover, in K cells, AMPK suppression using the small molecule inhibitor dorsomorphin (compound C) or shRNAs targeting AMPK α 1 or AMPK α 2 increased each of the metabolic and gene expression parameters and upregulated global DNA methylation levels, mimicking the effects of LKB1 loss (Extended Data Fig. 7g–q). Additionally, AMPK silencing blocked the effects of LKB1 re-expression on the growth and metabolic properties of KL cells (Extended Data Fig. 7r–u). Finally, KL cells were hypersensitive to mTOR inhibition using Torin 1, which suppressed the metabolic and transcriptional changes seen in these cells (Extended Data Fig. 7v–z). Thus, deregulation of AMPK–mTORC1 signalling contributes to the altered metabolic and epigenetic network in KL cells.

LKB1 loss silences retrotransposons

DNA methylation can contribute to transcriptional regulation and maintenance of genomic stability²⁴. We mapped methylation changes by conducting whole-genome bisulphite sequencing (WGBS) of K and KL cells (two independently derived lines per genotype). The data demonstrated a marked increase in mean CpG methylation in KL cells (Fig. 4a and Extended Data Fig. 8a). Comparison of methylation levels at non-repetitive 100-bp tiles revealed 3,395 hypermethylated and 1,270 hypomethylated regions in KL cells (FDR < 0.05 , methylation difference > 0.1 ; Fig. 4b and Extended Data Fig. 8b). However, the genes associated with differentially methylated regions had only modest overlap with those showing differential expression, suggesting that methylation of these elements may not prominently affect transcriptional regulation (Extended Data Fig. 8c). Thus, we focused on the repetitive portion of the genome. Notably, methylation in KL cells was particularly enriched at retrotransposon repeats (long terminal

repeats (LTRs), long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs)) compared to non-repeat elements (promoters, CpG islands and shores) (Fig. 4c, d and Extended Data Fig. 8d–k). These elements comprise a large proportion of the genome, have promoters that can be actively transcribed when unmethylated, and can influence regulation of host genes²⁵. Consistent with a functional role for the observed retrotransposon methylation, KL cells had a $> 50\%$ reduction in LINE-1 expression compared to K cells (Fig. 4e). Moreover, restoration of LKB1 suppressed LINE-1 methylation, as determined by 5mC–DNA immunoprecipitation–PCR analysis (meDIP–qPCR), and led to an approximately twofold upregulation of LINE-1 transcript expression (Fig. 4f, g). Importantly, PSAT1 knockdown also reduced LINE-1 methylation and induced its expression in KL cells (Fig. 4h, i), indicating that increased serine metabolism supports retrotransposon methylation and transcriptional silencing downstream of LKB1 loss.

Because retrotransposons can function as important modulators of host gene expression²⁵, we explored the correlation between differentially expressed genes in K versus KL cells and the presence of linked retrotransposons. Retrotransposon elements were significantly enriched (χ^2 test, $P < 0.0001$) in the gene bodies of differentially expressed genes as compared to the distribution of these elements across all genes (Extended Data Fig. 8l, m). By contrast, minimal enrichment was observed when promoter regions were considered. Such specificity is notable in light of evidence that intragenic methylation (including repeat methylation) can modify the activity of linked promoters and affect RNA processing^{25,26}. Thus, the LKB1–SGOC pathway alters the epigenetic landscape and dynamically regulates retrotransposon methylation and transcriptional activity, changes that appear to be associated with differences in host gene transcription.

KL cells are sensitive to DNMT inhibition

Consistent with the functional relevance of increased DNA methylation, KL cells were sensitive to shRNA-mediated knockdown of DNMT1 or DNMT3A, whereas K or KPC cells remained largely unaffected (Extended Data Fig. 9a–c). Similar specific sensitivity of KL cells was observed *in vivo* using doxycycline (dox)-inducible shRNAs to acutely deplete DNMT1 or DNMT3 (or shGFP control) once subcutaneous tumours reached 50 mm³ in size (Fig. 5a–c and Extended Data Fig. 9d–f). These data indicate that inhibition of DNA methylation is a potential vulnerability of KL cells. Indeed, KL cells were hypersensitive

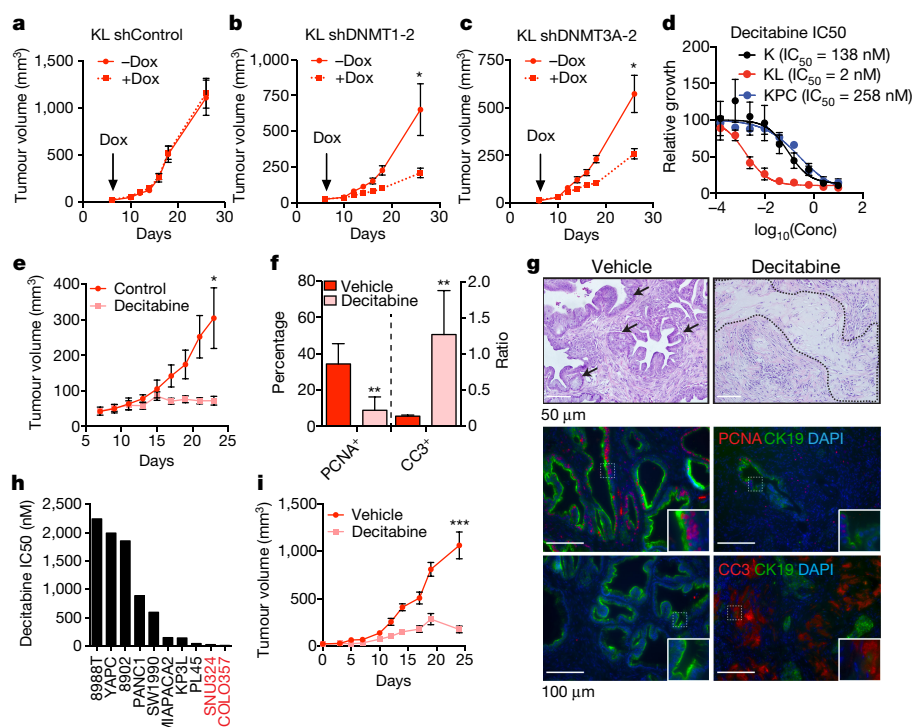


Figure 5 | LKB1 deficiency confers hypersensitivity to DNA methylation inhibitors. **a–c**, Volume of subcutaneous tumours derived from KL cells transduced with Dox-inducible shRNAs ($n = 6$ tumours per condition). **d**, Decitabine IC₅₀ for K, KL or KPC cells ($n = 4$). **e–g**, Decitabine treatment of KL xenografts ($n = 20$ tumours per condition). **e**, Tumour growth. **f**, Proportion of CK19⁺ tumour cells that are PCNA⁺ ($n = 4$ representative tumours), and ratio of cleaved-caspase 3 (CC3) to DAPI staining ($n = 4$ representative tumours). **g**, Haematoxylin and eosin

staining of representative tumours (top; arrows: tumour glands, dotted line: hyalinization), and staining for the indicated markers (quantified in **f**). Insets: threefold magnification. **h**, Decitabine IC₅₀ in human pancreatic cancer cell lines (red, LKB1 mutant; black, LKB1 wild type). **i**, Growth of COLO357 xenografts treated with decitabine (1 mg kg⁻¹) or vehicle ($n = 6$ per group). Data pooled from three (**h**) or representative of two (**d**) experiments. Error bars: s.e.m. (**a–c**, **i**), s.d. (**d–f**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

to the clinically approved DNMT inhibitor decitabine, as compared to K and KPC cell lines (IC₅₀ KL: 2 nM; K: 138 nM; KPC: 258 nM) (Fig. 5d and Extended Data Fig. 9g–i). Similar results were obtained using additional DNMT inhibitors (RG108, EGCG and SGI1027; Extended Data Fig. 9j–m). Importantly, decitabine treatment caused striking regression of subcutaneous KL tumours, associated with necrosis, hyalinization, replacement of tumour glands with fibrosis, decreased tumour cell proliferation, and pronounced apoptosis (Fig. 5e–g). By contrast, KPC xenografts were unaffected by decitabine (Extended Data Fig. 9n–q). Thus, LKB1 deficiency confers specific hypersensitivity to inhibition of DNA methylation *in vitro* and *in vivo*.

Available human LKB1 mutant pancreatic cancer cell lines (COLO357 and SNU324) showed similar vulnerabilities. They exhibited 70–90% inhibition of proliferation following PSAT1 knockdown, whereas a set of LKB1 wild-type pancreatic cancer lines showed modest responses (Extended Data Fig. 10a), consistent with prior results²⁷ in LKB1 wild-type lines. PSAT1 inactivation also inhibited the growth of LKB1 mutant COLO357 xenografts but not LKB1 wild-type MIAPACA2 xenografts (Extended Data Fig. 10b). Moreover, LKB1 restoration or PSAT1 silencing decreased global methylation in COLO357 cells but not in PANC1 or PATU-8988T cells (Extended Data Fig. 10c, d). LKB1 mutant lines also exhibited an increased response to decitabine *in vitro* and *in vivo*, and to the methionine salvage pathway inhibitors 3DZA and cycloleucine (Fig. 5h, i and Extended Data Fig. 10e). Thus, both mouse and human LKB1 mutant pancreatic tumour cells are sensitized to inhibition of the serine pathway and to DNMT inhibition.

Discussion

In summary, we have defined a metabolic state central to tumorigenesis resulting from LKB1 loss in which glucose and glutamine-derived

intermediates are channelled towards the SGOC network leading to increased DNA methylation (Extended Data Fig. 10f). Such interplay between metabolic control and epigenetic reprogramming has been proposed as a key mechanism for cancer development in the context of mutations in metabolic enzymes^{1,2,5}. The present work provides evidence for a broader role of metabolic and epigenetic crosstalk in cancer pathogenesis, revealing that LKB1 mutant pancreatic cancer cells have a marked dependency on pathways linking glycolysis, serine metabolism and DNA methylation.

It appears probable that the influence of aberrant metabolism on epigenetics contributes to cancer in the context of other oncogenic mutations. For instance, PI3K/AKT signalling controls acetyl-coenzyme A metabolism to support histone acetylation and regulate growth-promoting genes²⁸. Numerous oncogenic pathways rewire metabolism, and the circuits that are activated vary widely, with different genetic lesions favouring distinct fates of glucose, glutamine, fatty acids, and other nutrients. Thus, the resulting alterations in the levels of metabolites affecting chromatin regulation (for example, NAD⁺/NADH, FAD, O-linked *N*-acetylglucosamine, free fatty acids, SAM and acetyl-CoA) may be fundamental to cancer-promotion. Varying conditions in the tumour microenvironment may alter tumour phenotypes via similar processes^{29,30}.

The gain in DNA methylation at retrotransposons that we observed in KL cells is notable in light of the emerging view that these abundant repeat elements serve critical roles in host gene regulation²⁵. Moreover, there is increasing evidence that reactivation of silenced retrotransposons may underlie the therapeutic benefit of DNMT inhibitors, serving to induce a type I interferon anti-viral defence program^{31,32}. The basis for the widespread differences in response to DNMT inhibitors observed *in vitro* and in the clinic remains unclear. In this regard, our studies in KRAS mutant pancreatic ductal cells suggest that LKB1 status is a

genetic marker for DNMT inhibitor responsiveness. Because repeat element methylation does not strictly correlate with sensitivity to these treatments³³, the precise mechanisms of the specific response of KL cells to DNMT inhibitors remain to be defined. In this setting, it will be of interest to determine the relative contributions of IFN pathway activation and of modulation of host gene expression. Together, our data point to novel therapeutic vulnerabilities in the context of LKB1 mutations, suggesting that it might be possible to use agents targeting nodes of this network in defined patient subsets, although additional studies will be needed to broadly establish these associations in human cancers.

Our findings and other recent reports highlight the importance in cancer of the deregulation of serine biosynthesis, which can result from amplification of genes encoding pathway enzymes^{23,34} or their transcriptional activation downstream of oncogenic drivers, including NRF2³⁵, c-MYC^{36–38}, and mTOR³⁹. Serine metabolism can fuel diverse biosynthetic pathways and contribute to energy production. Correspondingly, its functional role in aberrant growth appears context dependent, relating variously to supporting nucleotide biosynthesis, NADPH production, TCA cycle anaplerosis, and DNA or histone methylation. These findings underscore the emerging interest in developing new approaches to targeting components of the SGOC network as cancer therapeutics.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 May 2015; accepted 27 September 2016.

Published online 31 October 2016.

- Gut, P. & Verdin, E. The nexus of chromatin regulation and intermediary metabolism. *Nature* **502**, 489–498 (2013).
- Etchegaray, J.-P. & Mostoslavsky, R. Interplay between metabolism and epigenetics: a nuclear adaptation to environmental changes. *Mol. Cell* **62**, 695–711 (2016).
- Carey, B. W., Finley, L. W. S., Cross, J. R., Allis, C. D. & Thompson, C. B. Intracellular α -ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **518**, 413–416 (2015).
- Shyh-Chang, N. *et al.* Influence of threonine metabolism on S-adenosylmethionine and histone methylation. *Science* **339**, 222–226 (2013).
- Losman, J.-A. & Kaelin, W. G. Jr. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev.* **27**, 836–852 (2013).
- Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
- Witkiewicz, A. K. *et al.* Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.* **6**, 6744 (2015).
- Su, G. H. *et al.* Germline and somatic mutations of the STK11/LKB1 Peutz-Jeghers gene in pancreatic and biliary cancers. *Am. J. Pathol.* **154**, 1835–1840 (1999).
- Giardiello, F. M. *et al.* Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology* **119**, 1447–1453 (2000).
- Korsse, S. E. *et al.* Pancreatic cancer risk in Peutz-Jeghers syndrome patients: a large cohort study and implications for surveillance. *J. Med. Genet.* **50**, 59–64 (2013).
- Ji, H. *et al.* LKB1 modulates lung cancer differentiation and metastasis. *Nature* **448**, 807–810 (2007).
- Chen, Z. *et al.* A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. *Nature* **483**, 613–617 (2012).
- Wingo, S. N. *et al.* Somatic LKB1 mutations promote cervical cancer progression. *PLoS One* **4**, e5137 (2009).
- Liu, W. *et al.* LKB1/STK11 inactivation leads to expansion of a prometastatic tumor subpopulation in melanoma. *Cancer Cell* **21**, 751–764 (2012).
- Shackelford, D. B. & Shaw, R. J. The LKB1-AMPK pathway: metabolism and growth control in tumour suppression. *Nat. Rev. Cancer* **9**, 563–575 (2009).
- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
- Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8**, 937–940 (2011).
- Mehrmohamadi, M., Liu, X., Shestov, A. A. & Locasale, J. W. Characterization of the usage of the serine metabolic network in human cancer. *Cell Reports* **9**, 1507–1519 (2014).
- Bardeesy, N. *et al.* Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl Acad. Sci. USA* **103**, 5947–5952 (2006).
- Locasale, J. W. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat. Rev. Cancer* **13**, 572–583 (2013).
- Mentch, S. J. *et al.* Histone methylation dynamics and gene regulation occur through the sensing of one-carbon metabolism. *Cell Metab.* **22**, 861–873 (2015).
- Maddocks, O. D. K., Labuschagne, C. F., Adams, P. D. & Vousden, K. H. Serine metabolism supports the methionine cycle and DNA/RNA methylation through *de novo* ATP synthesis in cancer cells. *Mol. Cell* **61**, 210–221 (2016).
- Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–350 (2011).
- Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
- Elbarbary, R. A., Lucas, B. A. & Maquat, L. E. Retrotransposons as regulators of gene expression. *Science* <http://dx.doi.org/10.1126/science.aac7247> (2016).
- Kulis, M., Queirós, A. C., Beekman, R. & Martín-Subero, J. I. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim. Biophys. Acta* **1829**, 1161–1174 (2013).
- Son, J. *et al.* Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature* **496**, 101–105 (2013).
- Lee, J. V. *et al.* Akt-dependent metabolic reprogramming regulates tumor cell histone acetylation. *Cell Metab.* **20**, 306–319 (2014).
- Shim, E.-H. *et al.* L-2-Hydroxyglutarate: an epigenetic modifier and putative oncometabolite in renal cancer. *Cancer Discov.* **4**, 1290–1298 (2014).
- Intlekofer, A. M. *et al.* Hypoxia induces production of L-2-hydroxyglutarate. *Cell Metab.* **22**, 304–311 (2015).
- Roulois, D. *et al.* DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* **162**, 961–973 (2015).
- Chiappinelli, K. B. *et al.* Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
- Treppendahl, M. B., Kristensen, L. S. & Grønbaek, K. Predicting response to epigenetic therapy. *J. Clin. Invest.* **124**, 47–55 (2014).
- Locasale, J. W. *et al.* Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nat. Genet.* **43**, 869–874 (2011).
- DeNicola, G. M. *et al.* NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nat. Genet.* **47**, 1475–1481 (2015).
- Sun, L. *et al.* cMyc-mediated activation of serine biosynthesis pathway is critical for cancer progression under nutrient deprivation conditions. *Cell Res.* **25**, 429–444 (2015).
- Ye, J. *et al.* Serine catabolism regulates mitochondrial redox control during hypoxia. *Cancer Discov.* **4**, 1406–1417 (2014).
- Zhang, W. C. *et al.* Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell* **148**, 259–272 (2012).
- Ben-Sahra, I., Howell, J. J., Asara, J. M. & Manning, B. D. Stimulation of *de novo* pyrimidine synthesis by growth signaling through mTOR and S6K1. *Science* **339**, 1323–1328 (2013).

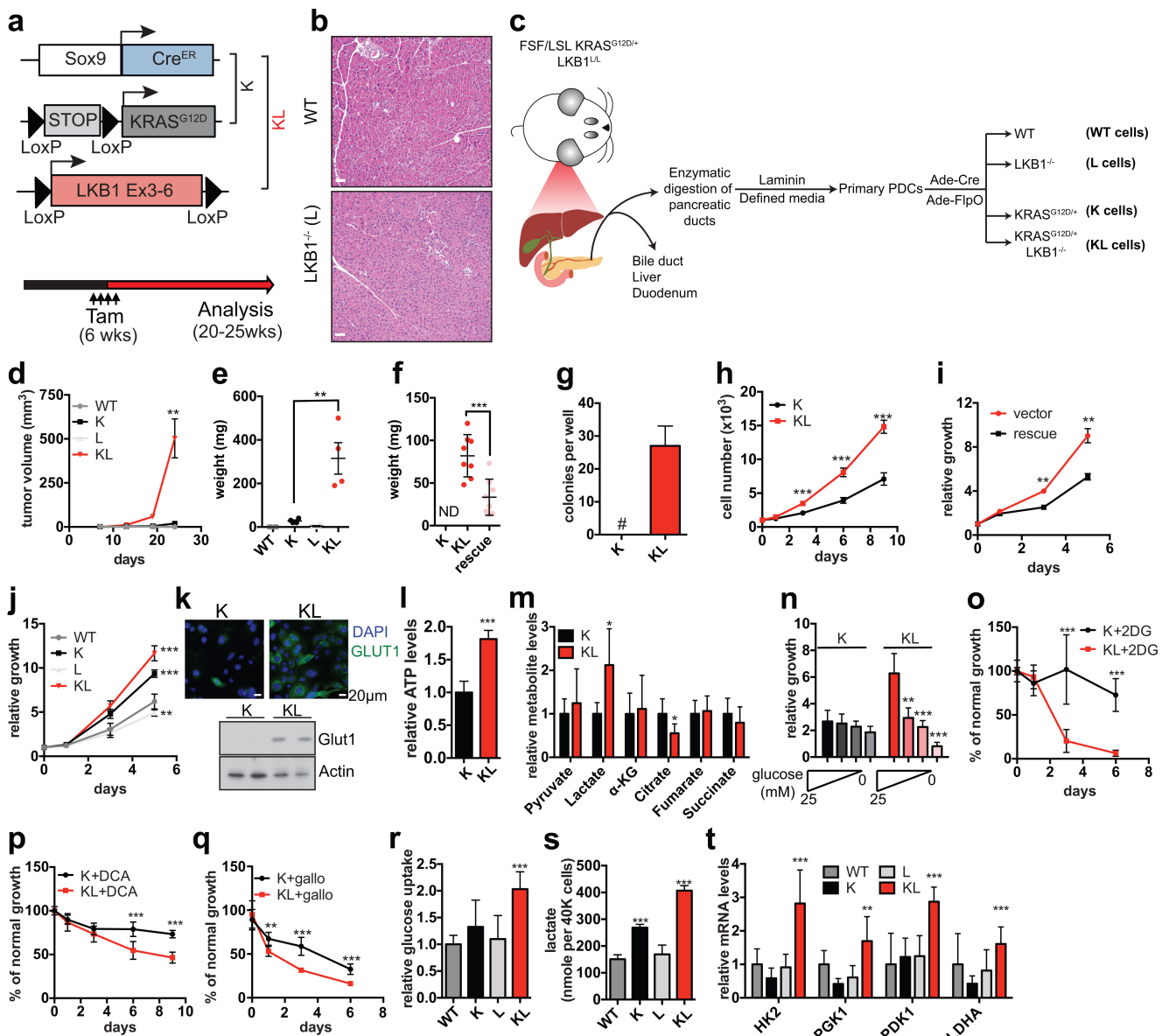
Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Kimmelman, K. Patra, L. J. Etchegaray, and R. Mostoslavsky for comments on the manuscript, and P. Foltopoulou, B. Martinez and Bardeesy laboratory members for advice. N.B. holds the Gallagher Endowed Chair in Gastrointestinal Cancer Research and received support from the Granara-Skerry Trust, the Linda J. Verville Foundation and the Begg Family, and grants from the NIH (P01 CA117969-07, R01 CA133557-05). F.K. is supported by a Hirshberg Foundation Career Development Award. F.K. and N.B. were supported by NIH grant P50CA127003 and are members of the Andrew Warshaw Institute.

Author Contributions F.K. and N.B. conceived and designed the study. F.K., A.R. and J.M.N. performed cell-based and mouse experiments. T.C. assisted with mouse experiments. F.K. and B.N.N. performed and interpreted the tracing experiments. M.B. and W.H. performed proteomics. F.K. and M.L. performed the OCR measurements. M.C.H. and D.N.H. provided essential samples and data analysis. Y.Y.L. performed computational analysis. H.G. prepared WGBS libraries. R.K. and A.M. analysed and interpreted the WGBS data. P.S.H., K.K.W., O.S.S. and N.J.D. assisted with data interpretation. F.K. and N.B. wrote the manuscript with feedback from all authors.

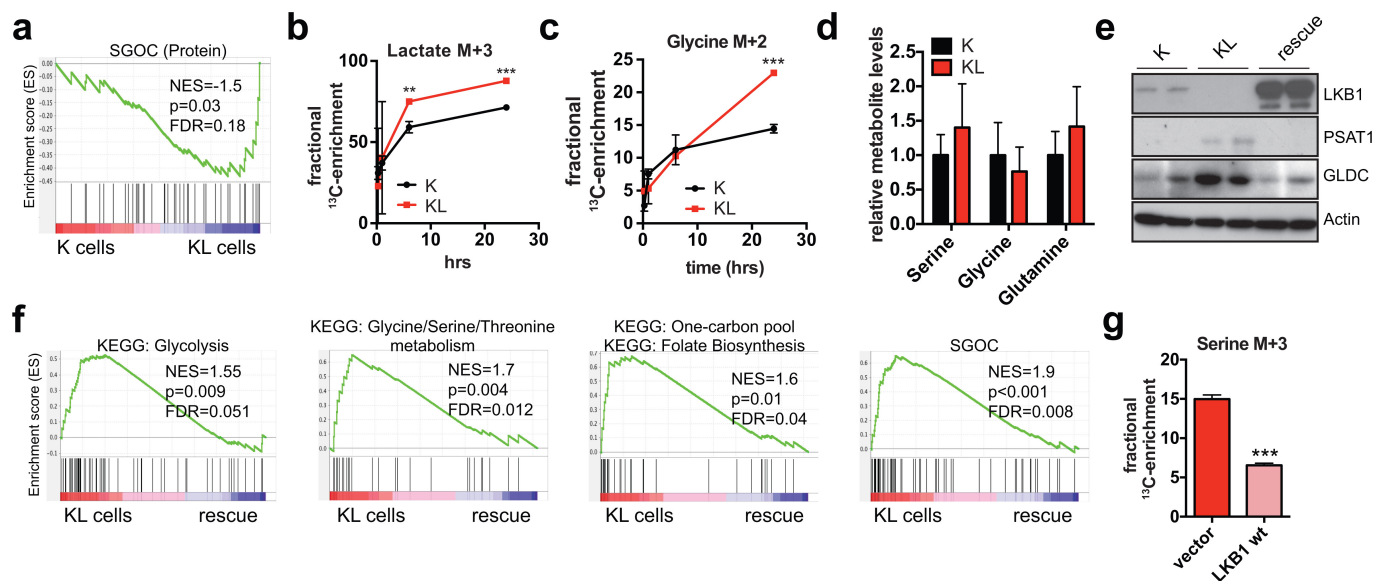
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.B. (bardeesy.nabeel@mgh.harvard.edu).

Reviewer Information *Nature* thanks M. Rehli and the other anonymous reviewer(s) for their contribution to the peer review of this work.



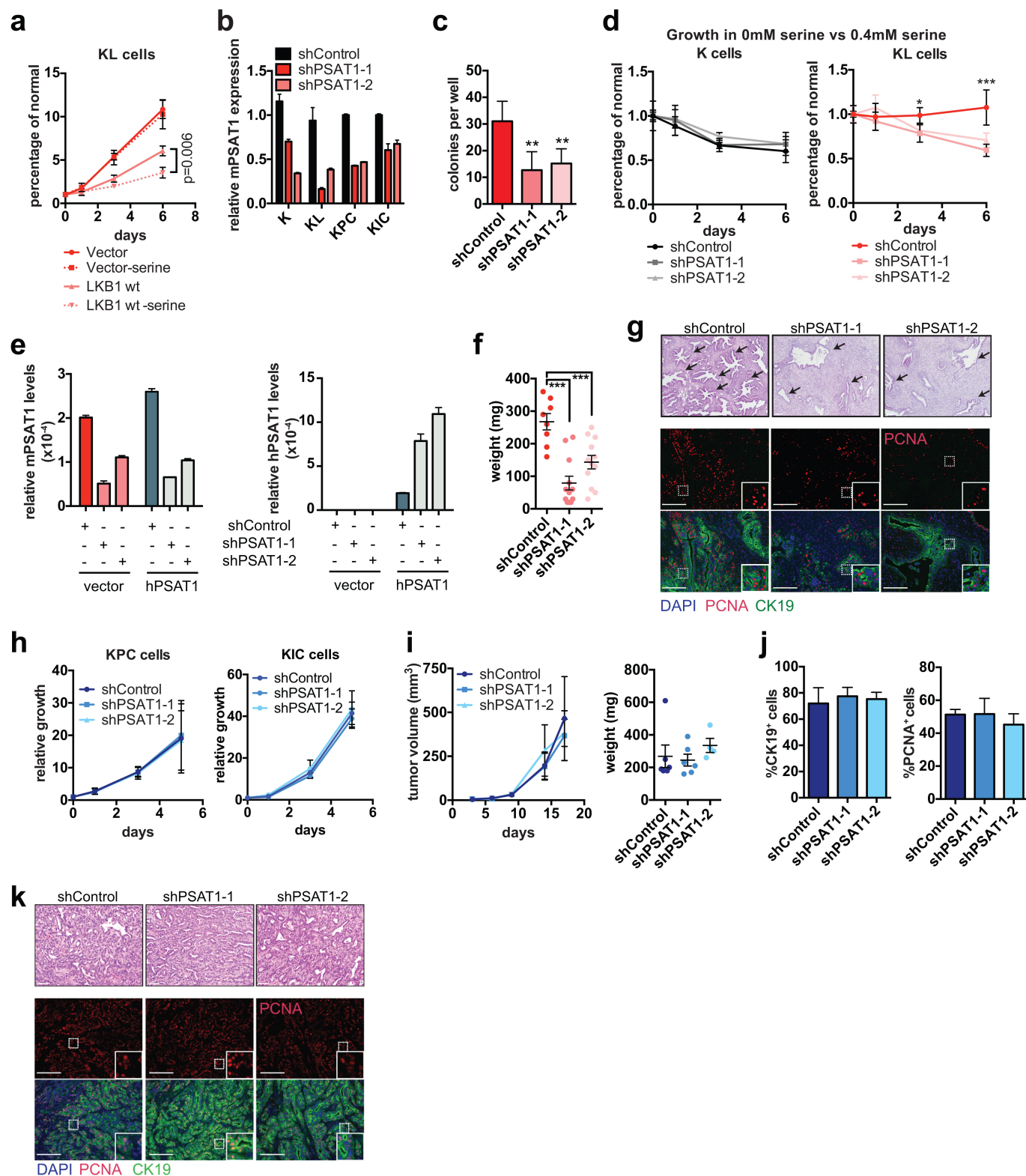
Extended Data Figure 1 | LKB1 suppresses KRAS^{G12D}-driven tumorigenesis and limits glycolysis in primary pancreatic ductal epithelial cells. **a**, Schematic of GEM models. Sox9-Cre^{ER}, LKB1^{L/L}, and LSL-KRAS^{G12D/+} mice were crossed to generate four cohorts: WT (Sox9-Cre^{ER}), L (Sox9-Cre^{ER};LKB1^{L/L}), K (Sox9-Cre^{ER};LSL-KRAS^{G12D/+}) and KL (Sox9-Cre^{ER};LSL-KRAS^{G12D/+};LKB1^{L/L}). Genetic lesions were induced by intraperitoneal injections of tamoxifen at 6 weeks of age, after which mice were observed for signs of disease and killed when KL animals were moribund (20–25 weeks of age). The WT, K, and L mice had no signs of illness or other abnormalities at this time point. **b**, Haematoxylin and eosin-stained sections of representative pancreata from WT and L mice ($n = 4$ mice per group). Scale bars, 50 μm . **c**, Schematic of primary pancreatic ductal epithelial cell system. Pancreatic ductal epithelial cells were isolated from LSL-KRAS^{G12D/+} and LSL-KRAS^{G12D/+};LKB1^{L/L} mice and infected with Adeno-Cre to generate K and KL cells. For studies comparing K, L, KL and WT genotypes, cells were isolated from FSF-KRAS^{G12D/+};LKB1^{L/L} mice and infected with Adeno-Flipase and/or Adeno-Cre, or neither. **d**, Volume and **e**, weight of subcutaneous tumours derived from ductal cells of the indicated genotypes ($n = 4$ tumours per group). Error bars show s.e.m. For source data on tumour volume, see Supplementary Data Table 4. **f**, Weight of subcutaneous tumours from K ($n = 6$), KL ($n = 8$) and KL cells transduced with retrovirus expressing LKB1 cDNA (rescue, $n = 8$). **g**, Number of colonies formed in soft agar by K (none detected) or KL cells ($n = 6$ independent biological replicates). Error bars show s.e.m. **h**, Proliferation of K and KL cells in nutrient-replete medium ($n = 4$).

i, Proliferation of KL cells transduced with retroviruses expressing empty vector or LKB1 (rescue; $n = 3$). **j**, Proliferation of wild-type (WT), KRAS^{G12D/+} (K), LKB1^{-/-} (L) and KRAS^{G12D/+};LKB1^{-/-} (KL) cells ($n = 6$). **k-t**, *In vitro* studies of K and KL cells. **k**, Detection of GLUT1 (SLC2A1) by immunofluorescence (scale bar, 20 μ m) or immunoblot. 2-(4-amidinophenyl)-1H-indole-6-carboxamide (DAPI) was used to visualize nuclei. Actin was used as the loading control. For gel source images see Supplementary Data Fig. 1. **l**, Steady-state ATP levels under nutrient-replete conditions measured by CellTiterGlo (Promega), normalized to cell number and expressed as relative to ATP levels in K cells ($n = 4$). **m**, Intracellular levels of pyruvate, lactate, and TCA cycle metabolites as detected by GC-MS. Values are normalized to cell number. Data are expressed as relative to the levels in K cells ($n = 6$ biological replicates). **n**, Three-day proliferation of cells in 25 mM glucose or acutely switched to media with the indicated reduced glucose concentrations ($n = 6$). Data are expressed as relative to day 0. **o-q**, Proliferation of cells treated with 5 mM 2-deoxyglucose (2DG) (**o**), 5 mM dichloroacetate (DCA) (**p**) or 20 μ M galloflavin (gallo) (**q**). Values are expressed as percentage of normal growth (2DG $n = 6$, DCA $n = 8$, gallo $n = 6$). **r-t**, WT, K, L and KL cells were measured for glucose uptake using 2NBDG (**r**, $n = 6$), lactate release into the medium (**s**, $n = 4$), and expression of glycolytic genes (**t**, $n = 6$). Data are pooled from two (**j**, **l**, **n-r**) or three (**t**) experiments or representative of two (**h**, **i**, **s**) experiments. For all panels, error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



Extended Data Figure 2 | LKB1 loss induces the serine-glycine-one-carbon network. **a**, GSEA showing enrichment of proteins involved in the serine-glycine-one-carbon network¹⁸ in KL cells compared to K cells using global proteomics ($n = 4$ samples per group). **b**, Plot of isotopomer abundance of U^{13}C glucose-derived intracellular M+3 lactate over time ($n = 3$ independent biological replicates). **c**, Plot of isotopomer abundance of U^{13}C glucose-derived intracellular M+2 glycine over time ($n = 3$ independent biological replicates). **d**, Intracellular levels of serine, glycine and glutamine as detected by GC-MS. Values normalized to cell number. Data are expressed as relative to the levels in K cells ($n = 6$ biological replicates). **e**, PSAT1 and GLDC expression determined by immunoblot

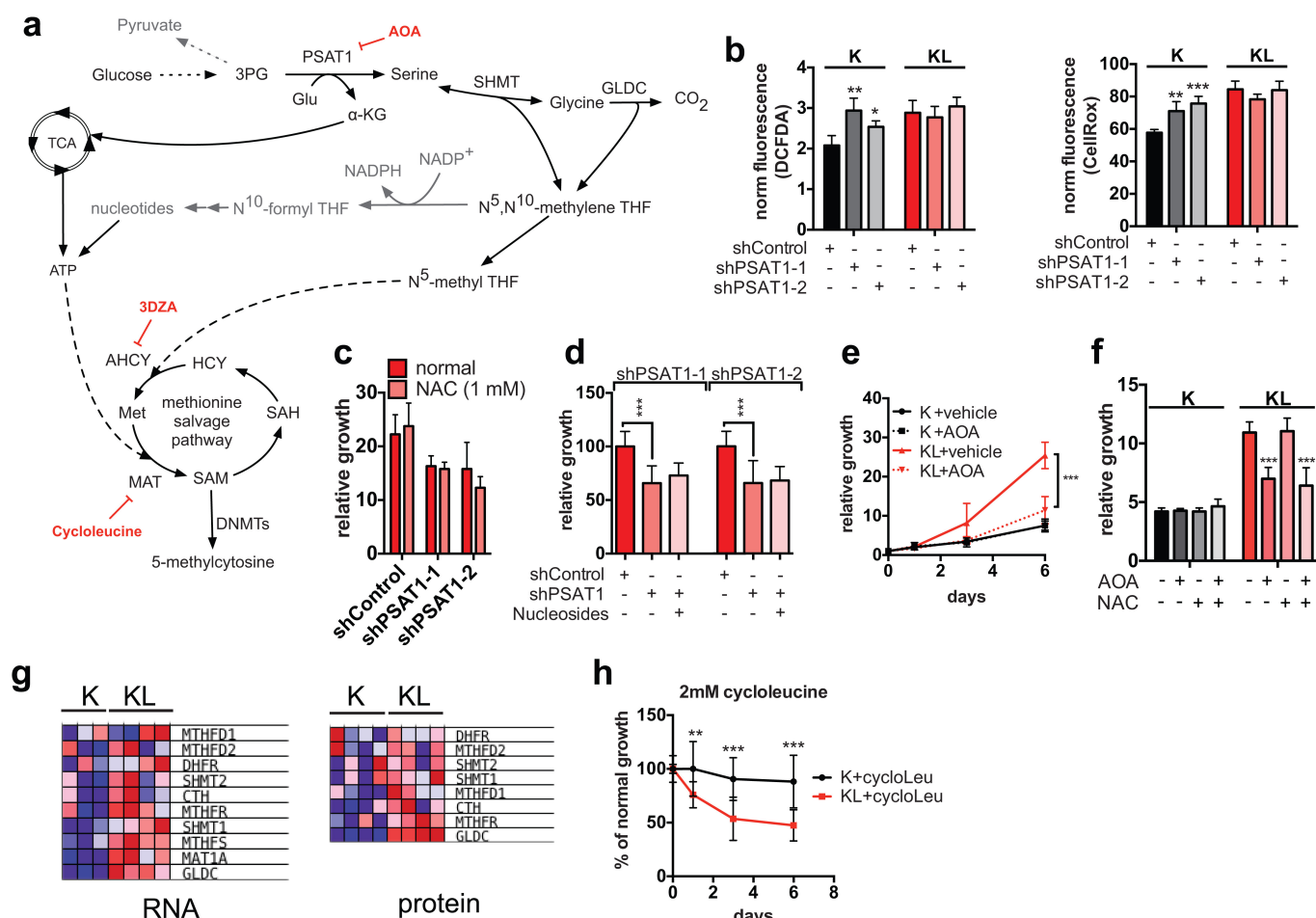
in K and KL cells and in KL cells transduced with LKB1 cDNA. Actin was used as loading control. For gel source images see Supplementary Data Fig. 1. **f**, GSEA of RNA-seq data showing suppression of genes involved in glycolysis, serine biosynthesis, folate cycle and the serine-glycine-one-carbon network upon re-expression of LKB1 cDNA in KL cells (rescue, $n = 2$ samples) compared to parental KL cells ($n = 4$ samples). **g**, Plot of isotopomer abundance of U^{13}C glucose-derived intracellular M+3 serine 6 h after addition of U^{13}C glucose ($n = 3$ independent biological replicates). For all panels, error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



Extended Data Figure 3 | See next page for caption.

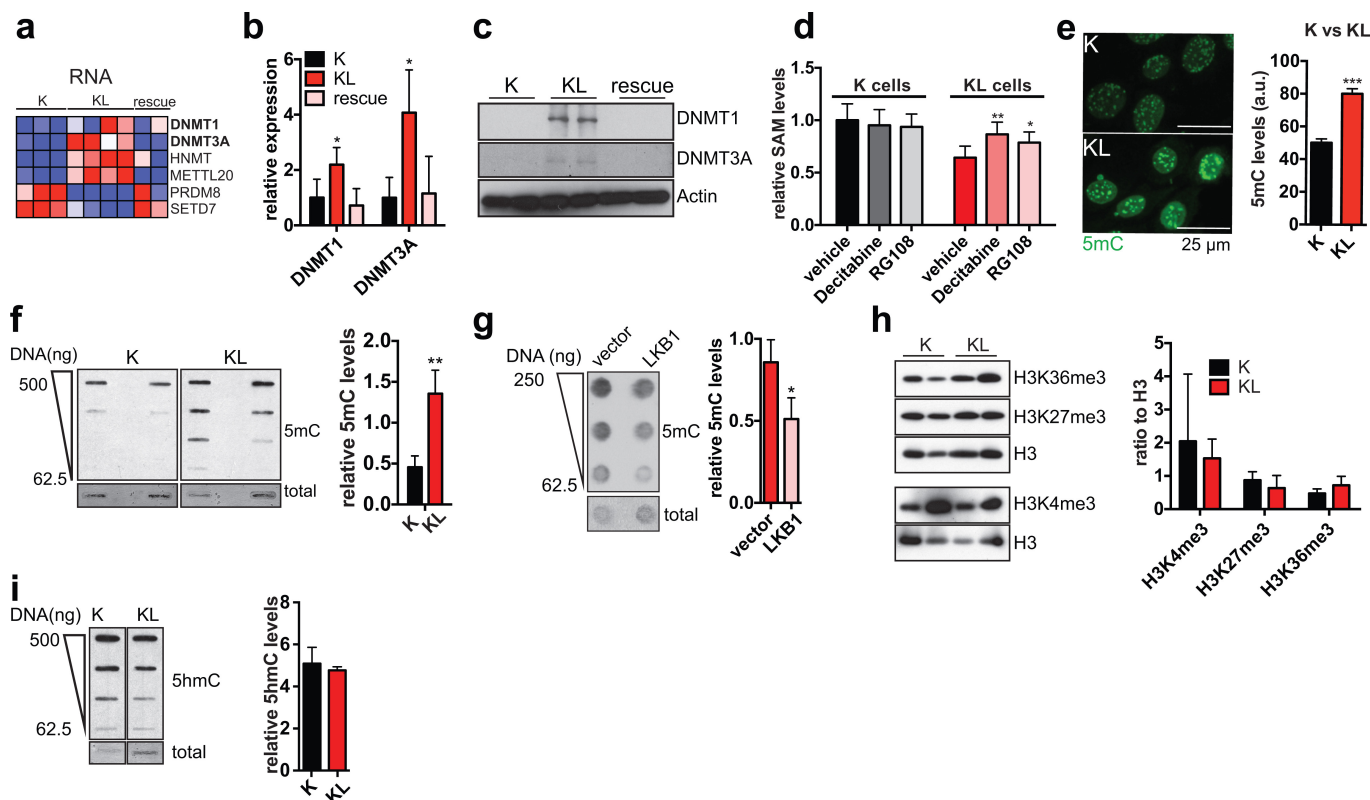
Extended Data Figure 3 | The *de novo* serine biosynthesis pathway is required for KL but not KPC tumour growth. **a**, Proliferation of KL cells transduced with vector or LKB1 cDNA and cultured in the presence or absence of 0.4 mM serine. Growth is expressed as relative to day 0 ($n = 3$). LKB1 re-expression slows growth and results in sensitivity to serine deprivation. **b**, qRT-PCR showing effective knockdown of PSAT1 in K, KL, KPC and KIC cells transduced with shControl or two different shRNAs against PSAT1 ($n = 2$). Data are expressed as relative to shControl for each cell line. 18S rRNA was used for normalization. **c**, Number of colonies formed in soft agar by KL cells transduced with shControl or two different shRNAs against PSAT1 ($n = 6$). **d**, Proliferation of K or KL cells transduced with shControl or two independent shRNAs against PSAT1 in the absence of serine. Data are expressed as percentage of growth in the presence of 0.4 mM serine ($n = 6$). **e**, qRT-PCR showing effective knockdown of endogenous mouse PSAT1 (mPSAT1, endogenous) and forced expression of human PSAT1 (hPSAT1, exogenous) in KL cells transduced with shControl, shPSAT1-1, shPSAT1-2, vector or hPSAT1 cDNA. Data are normalized to 18S rRNA ($n = 2$). **f**, Weight at the time of harvesting of subcutaneous tumours from KL cells transduced with shControl ($n = 8$), shPSAT1-1 ($n = 12$), or shSPAT1-2 ($n = 12$). Error bars show s.e.m. **g**, Haematoxylin and eosin-stained sections and immunofluorescence analysis of representative tumours derived from subcutaneous injections of KL cells transduced with shControl,

shPSAT1-1, or shPSAT1-2. Note that PSAT1 knockdown tumours have a reduction in malignant glands (arrows) relative to the fibrotic stroma. Lower panels: anti-CK19 (green) was used to visualize the neoplastic epithelium and anti-PCNA (red) was used to mark proliferation. DAPI was used to stain nuclei (blue). **h**, Proliferation of KPC and KIC pancreatic cancer cells transduced with shControl or two independent shRNAs against PSAT1. Growth is expressed as relative to day 0 ($n = 6$ for KPC and $n = 4$ for KIC). **i–k**, KPC cells were transduced with shControl ($n = 6$), shPSAT1-1 ($n = 6$) or shSPAT1-2 ($n = 4$) and injected into SCID mice. **i**, Volume (left) and weight (right) of subcutaneous tumours. Error bars show s.e.m. For source data on tumour volume, see Supplementary Data Table 4. **j**, **k**, Tumours in **i** were stained using anti-CK19 antibody (green) to visualize the neoplastic epithelium and anti-PCNA staining (red) to mark proliferating cells. DAPI was used to stain nuclei (blue). The proportion of stained CK19⁺ cells is quantified in **j** ($n = 4$, representative tumours) and CK19⁺ cells with nuclear PCNA staining are quantified in **j**. There are no significant effects on any of these parameters. **k**, Haematoxylin and eosin-stained sections (top) and immunofluorescence analysis (bottom) of representative tumours. Scale bars, 100 μ m. Insets show threefold magnification. Data pooled from two (**c**, **d**) or representative of two (**a**, **b**, **e**) or three (**h**) experiments. For all panels, error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



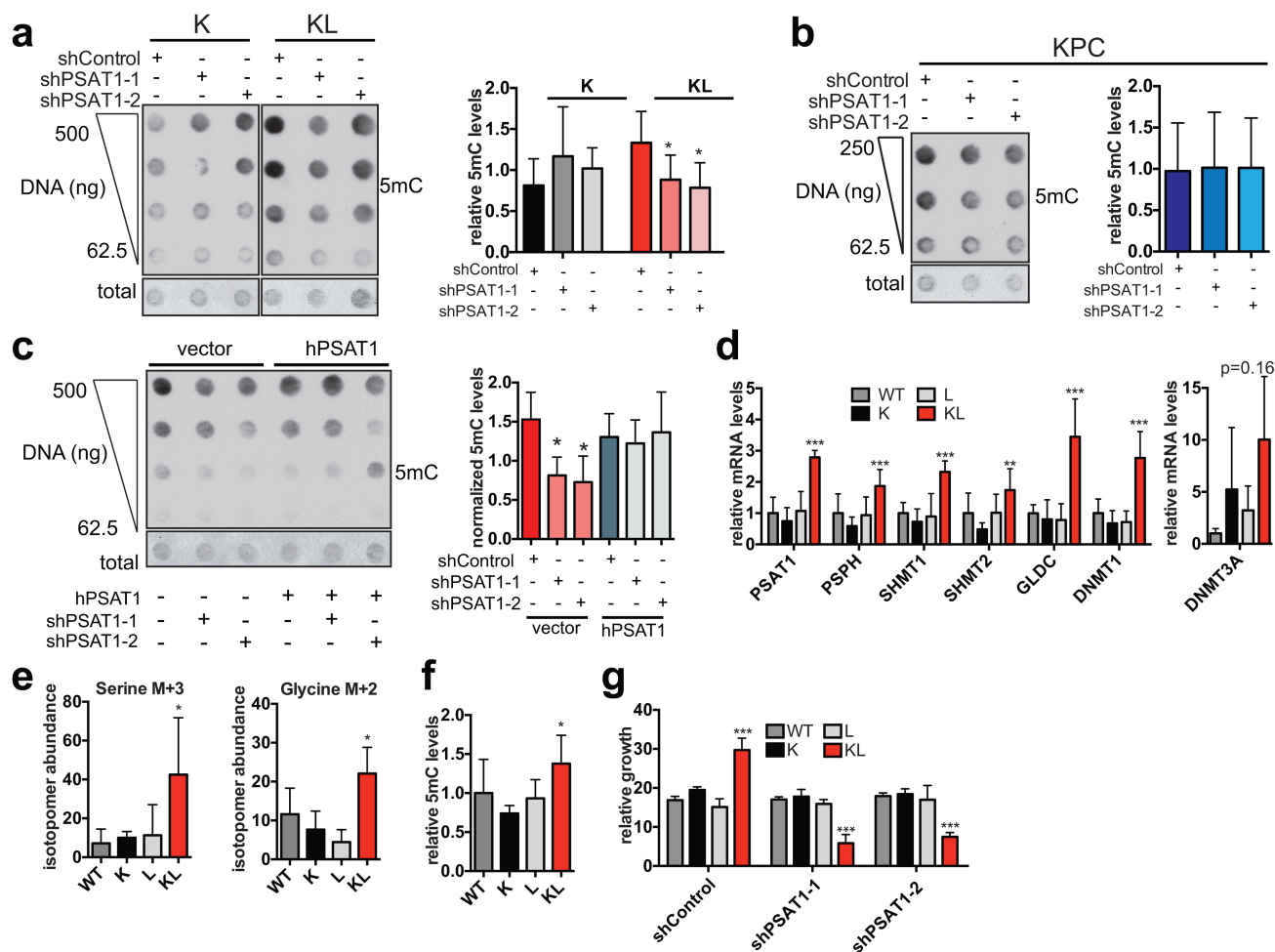
Extended Data Figure 4 | Characterization of serine–glycine–one-carbon pathway in KL cells. **a**, Detailed graph of SGOC network. Enzymatic inhibitors used in this study are marked in red. **b**, ROS in K or KL cells transduced with shControl, shPSAT1-1 or shPSAT1-2 measured by DCFDA (left) and CellROX staining (right). Data are normalized to cell number ($n = 4$). **c**, Six-day proliferation assay of KL cells transduced with shControl, shPSAT1-1 or shPSAT1-2, showing lack of growth rescue by N-acetylcysteine (NAC). Data are expressed as relative to day 0 ($n = 6$). **d**, Three-day growth assay of KL cells transduced with shControl, shPSAT1-1 or shPSAT1-2, showing the lack of rescue by excess nucleosides (adenosine, guanosine, thymidine, uridine, cytidine; 1 mM each). Data are presented as percentage of the growth of shControl cells

($n = 16$). **e**, Proliferation of K or KL cells treated with aminooxyacetate (AOA). Data are expressed as relative to day 0 ($n = 8$). **f**, Five-day proliferation of K or KL cells treated with AOA and/or NAC. Data are expressed as relative to day 0 ($n = 6$). **g**, Data from RNA-seq (left) and quantitative proteomics (right) showing levels of genes involved in the production of SAM in K and KL. The data plotted are expressed as mean-centred values. **h**, Proliferation of K or KL cells treated with 2 mM cycloleucine. Data are expressed as percentage of the growth of vehicle treated cells ($n = 12$). Data are pooled from two (**b–f**, **h**) experiments. For all panels, error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



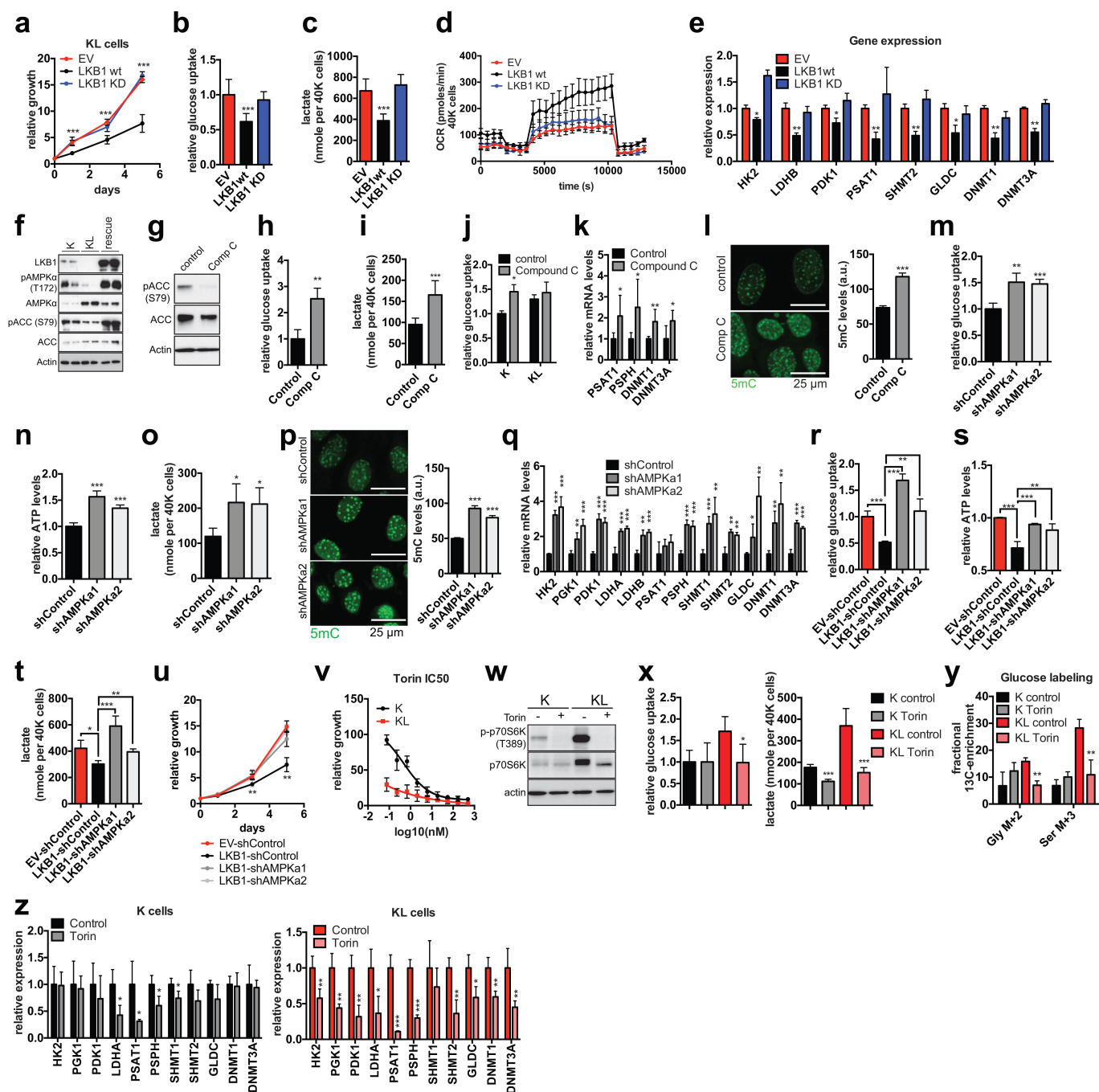
Extended Data Figure 5 | Deletion of LKB1 induces DNMT1 and DNMT3A expression and increases global DNA methylation. **a**, Heat map of RNA-seq data showing levels of the differentially regulated SAM-using enzymes. Plotted data are expressed as mean centred values. **b**, Expression of DNMT1 and DNMT3A in K, KL, and KL cells transduced with LKB1 cDNA (rescue) was measured by qRT-PCR. Levels were normalized to 18S rRNA. Data are expressed as relative to K cells ($n=4$, representative of two experiments). **c**, Immunoblots of lysates from K, KL or rescue cells were probed for DNMT1 or DNMT3A. Actin was used as loading control. For gel source images see Supplementary Data Fig. 1. **d**, Measurement of SAM in K and KL cells treated with 5-aza-2-deoxycytidine (decitabine) or RG108 for 3 days. In each case, data are expressed as relative to the amount of SAM in vehicle-treated K cells, which was arbitrarily set to 1 ($n=6$ independent replicates). **e**, Immunofluorescence staining and quantitation of 5mC in K or KL cells

(77–130 cells). Scale bar, 25 μ m. **f**, Dot blot of DNA isolated from K or KL cells probed with anti-5mC antibody. Quantified signal was normalized to total DNA as measured by methylene blue staining ($n=4$, independent replicates). **g**, Dot blot of DNA isolated from KL cells transduced with empty vector or LKB1 cDNA probed with anti-5mC antibody. Quantified signal was normalized to total DNA as measured by methylene blue staining ($n=3$, independent replicates). **h**, Immunoblot analysis of histone 3 (H3) methyl marks from K or KL cells. Data are normalized to total H3 (K4me3, $n=2$; K27me3, $n=5$; K36me3, $n=5$, independent replicates). For gel source images see Supplementary Data Fig. 1. **i**, Dot blot of DNA isolated from K or KL cells probed with anti-5hmC antibody. Quantified signal was normalized to total DNA as measured by methylene blue staining ($n=4$, independent replicates). For all panels, error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



Extended Data Figure 6 | Serine pathway activity sustains DNA methylation in KL cells. **a**, Dot blot of DNA isolated from K or KL cells transduced with shControl, shPSAT1-1 or shPSAT1-2 probed with anti-5mC antibody. Total DNA was visualized by methylene blue staining. Graph shows quantified signal normalized to total DNA as measured by methylene blue staining (K cells, $n=4$; KL cells, $n=8$, independent replicates). **b**, Dot blot of DNA isolated from KPC cells transduced with shControl, shPSAT1-1 or shPSAT1-2 probed with anti-5mC antibody. The graph shows quantified signal normalized to total DNA ($n=4$, independent replicates). **c**, Dot blot of DNA probed with anti-5mC antibody. DNA was isolated from KL cells first transduced with vector or human PSAT1 then transduced with shControl, shPSAT1-1 or shPSAT1-2. Graph shows quantified signal normalized to total DNA as measured by

methylene blue staining ($n=3$, independent replicates). **d**, Expression of serine pathway genes and DNMT genes in WT, K, L or KL cells by qRT-PCR. Data are normalized to 18S and expressed as relative to K cells ($n=6$, pooled data from two experiments). **e**, Plots of isotopomer abundance of U[^{13}C]glucose-derived M+3 serine and M+2 glycine, 6 h after addition of U[^{13}C]glucose ($n=6$ independent biological replicates). **f**, Quantified DNA dot blot signal of DNA isolated from WT, K, L or KL cells probed with anti-5mC antibody normalized to total DNA as measured by methylene blue staining ($n=4$, independent replicates). **g**, Five-day growth of WT, K, L or KL cells transduced with shControl or two shRNAs against PSAT1. Data are expressed as relative to day 0 ($n=4$, pooled from two experiments). For all panels, error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

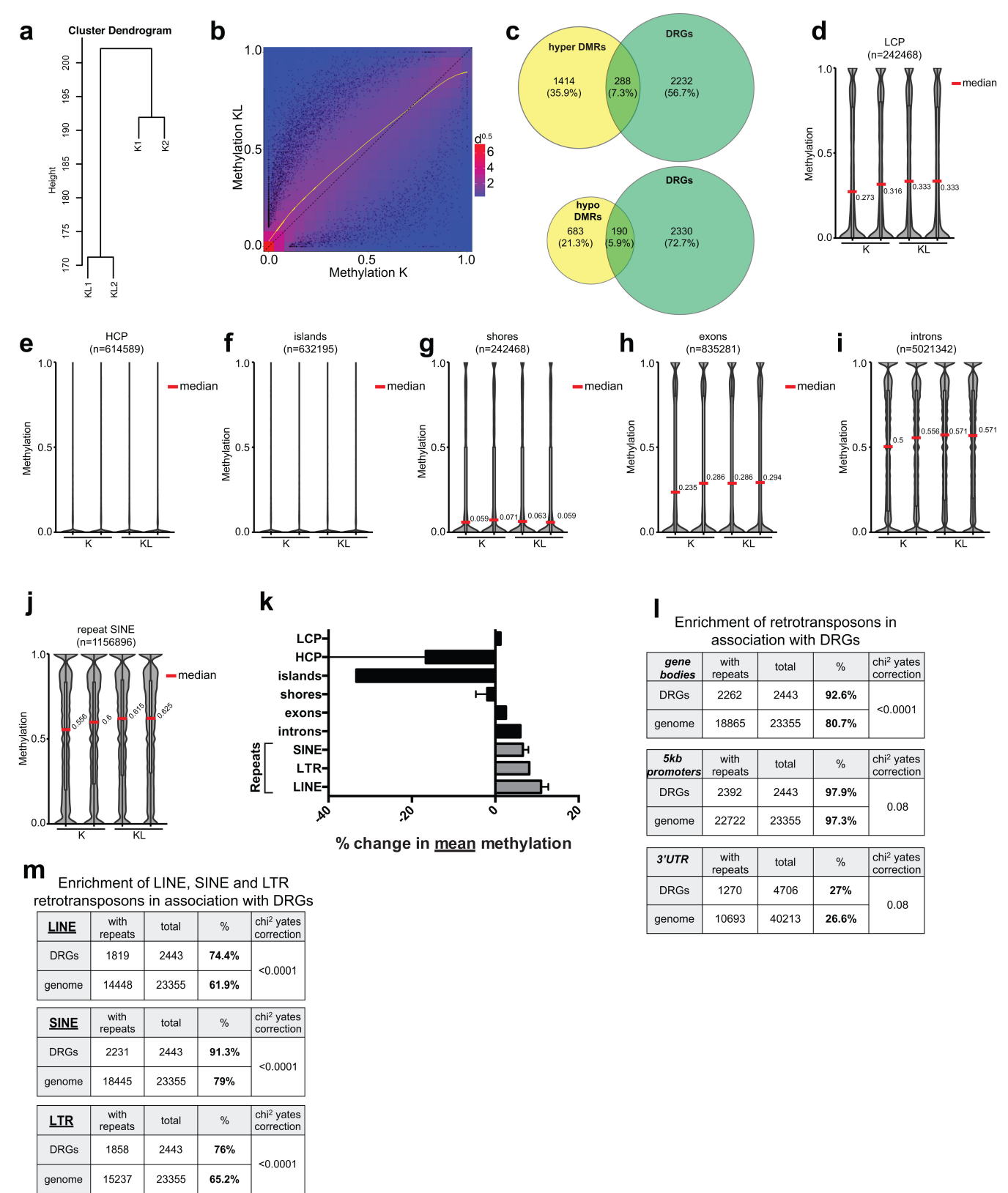


Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | LKB1-mediated regulation of glycolysis–SGOC–DNMT pathway involves the AMPK–mTOR axis.

a–e, KL cells were transduced with vector, wild-type LKB1 or kinase-dead LKB1. **a**, Proliferation, expressed as relative to day 0 ($n = 4$). **b**, Glucose uptake measured using 2NBDG followed by fluorimetry (data normalized to cell number and expressed as relative to KL-vector cells ($n = 8$)). **c**, Lactate levels measured by fluorimetry 3 h after medium change, normalized to cell number ($n = 8$). **d**, Oxygen consumption rates measured in normal duct medium, followed by injections with 4 μ M oligomycin (O), 4 μ M FCCP (F), or 4 μ M antimycin A (A) ($n = 3$). **e**, Expression of the indicated genes measured by qRT-PCR ($n = 4$), with levels normalized to 18S rRNA. Data are expressed as relative to KL-empty vector (EV) cells. **f**, Immunoblot of K, KL or KL cells transduced with LKB1 cDNA (rescue). For gel source images see Supplementary Data Fig. 1. **g**, Immunoblot of K cells treated overnight with 10 μ M Compound C. For gel source images see Supplementary Data Fig. 1. **h**, **i**, Glucose uptake (**h**) and 3 h lactate production (**i**) in the same cells as in **h**. Data are normalized to cell number and expressed as relative to control K cells (glucose uptake, $n = 3$; lactate, $n = 4$). **j**, Glucose uptake in K or KL cells treated with vehicle or 10 μ M Compound C. Data are normalized to cell number and expressed as relative to control K cells ($n = 4$). Note the blunted response of KL cells. **k**, qRT-PCR analysis of the indicated genes in the same cells as in **h**. Levels are normalized to 18S rRNA. Data are expressed as relative to control K cells ($n = 6$). **l**, Immunostaining for 5-mC (left) and quantification of staining (right) in K cells treated with vehicle or Compound C for 4 days (158–163 cells). **m–q**, K cells were transduced with shControl or shRNAs against AMPK α 1 or AMPK α 2. **m**, Glucose uptake. Data are normalized to cell number and expressed as relative to shControl-treated cells ($n = 4$).

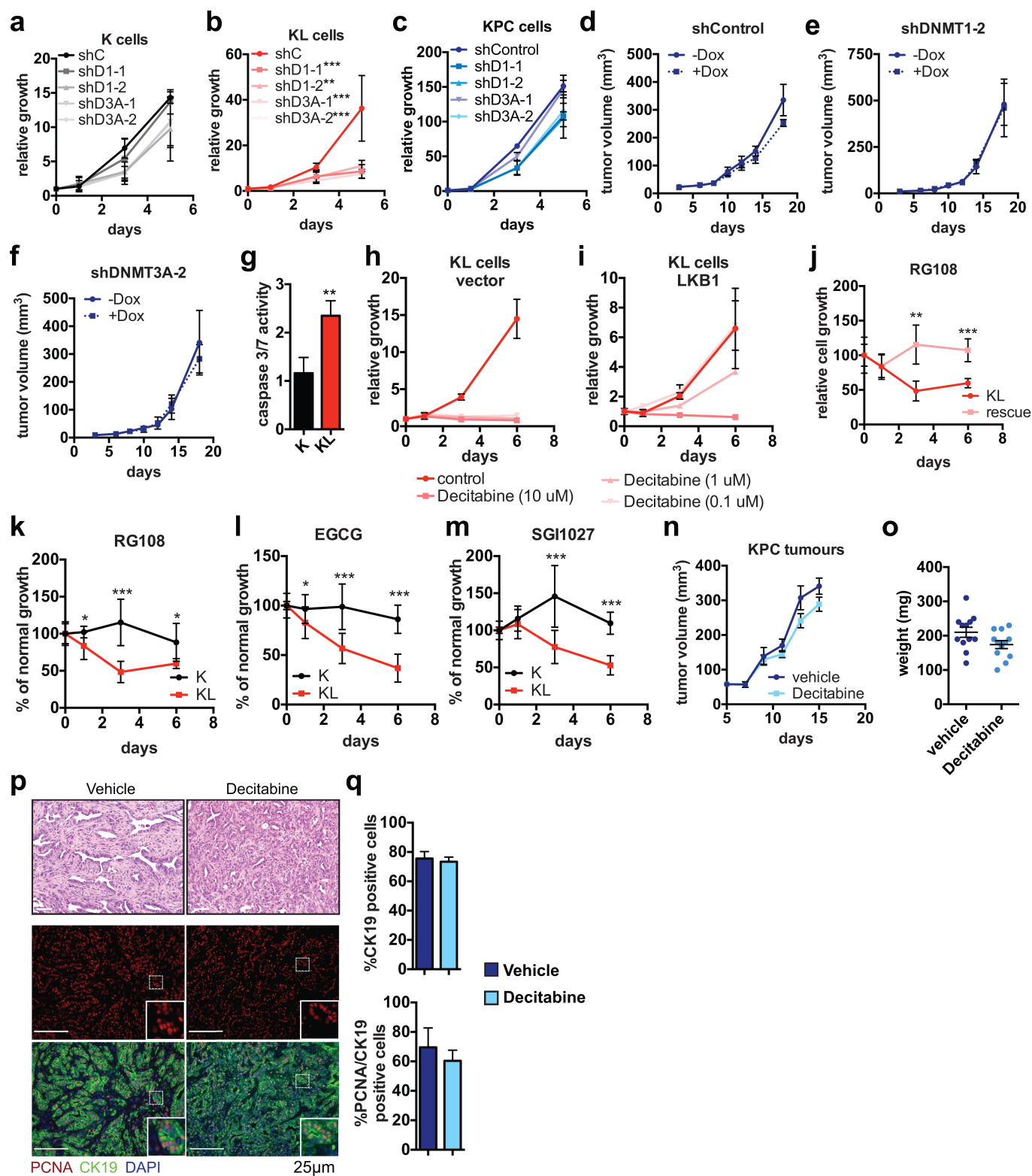
n, Steady-state ATP levels measured with CellTiterGlo (Promega), normalized to cell number and expressed as relative to ATP levels in shControl-treated cells ($n = 4$). **o**, Lactate levels measured by fluorimetry 3 h after medium change. Data are normalized to cell number ($n = 3$). **p**, Immunostaining for 5-mC (left) and quantification of staining (right) (159–296 cells). **q**, Expression of the indicated genes determined by qRT-PCR. Levels are normalized to 18S rRNA. Data are expressed as relative to shControl cells ($n = 4$). **r–u**, KL cells were transduced with vector or an LKB1 cDNA and shControl or shRNAs against AMPK α 1 or AMPK α 2. **r**, Glucose uptake. Data are normalized to cell number and expressed as relative to vector-shControl cells ($n = 4$). **s**, Steady-state ATP levels normalized to cell number and expressed as relative to ATP levels in EV-shControl cells ($n = 4$). **t**, Lactate levels 3 h after medium change. Data are normalized to cell number ($n = 4$). **u**, Proliferation expressed as relative to day 0 ($n = 3$). **v**, Impact of Torin 1 treatment on growth of K and KL cells ($n = 4$). **w**, Immunoblot of K and KL cells treated with vehicle or 25 nM Torin 1. For gel source images see Supplementary Data Fig. 1. **x**, Glucose uptake and 3 h lactate production in K or KL cells treated overnight with 25 nM Torin 1. Data are normalized to cell number ($n = 4$). **y**, Isotopomer abundance of U 13 C]glucose-derived serine and glycine in the same cells as in **r** ($n = 3$ independent replicates). Cells were labelled with U 13 C]glucose for 6 h. **z**, Expression of the indicated genes in Torin-treated K (left) or KL cells (right) determined by qRT-PCR. Levels are normalized to 18S rRNA. Data are expressed as relative to control-treated cells ($n = 4$). Data are pooled from two (**b**, **c**, **e**, **i–n**, **p–t**, **v**, **x**, **z**) or representative of two (**a**, **d**, **h**, **o**) or three (**u**) experiments. Error bars show s.e.m. in **d**, s.d. in all other panels. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Loss of LKB1 increases DNA methylation in retrotransposon elements. **a**, Hierarchical clustering of K and KL samples based on methylation levels at 100-bp autosomal genomic tiles as measured by whole-genome bisulfite sequencing. **b**, Two-dimensional density plot for methylation at non-repetitive 100-bp autosomal genomic tiles in KL versus K cells. The yellow line shows the mean methylation of tiles. The dots represent differentially methylated tiles (FDR q -value < 0.05 , methylation change > 0.1). There are 3,395 hypermethylated tiles and 1,270 hypomethylated tiles in KL cells versus K cells. **c**, Overlap of genes associated with differentially methylated regions and differentially regulated genes (hyper DMRs, hypermethylated regions; hypo DMRs, hypomethylated regions; DRGs, differentially regulated genes).

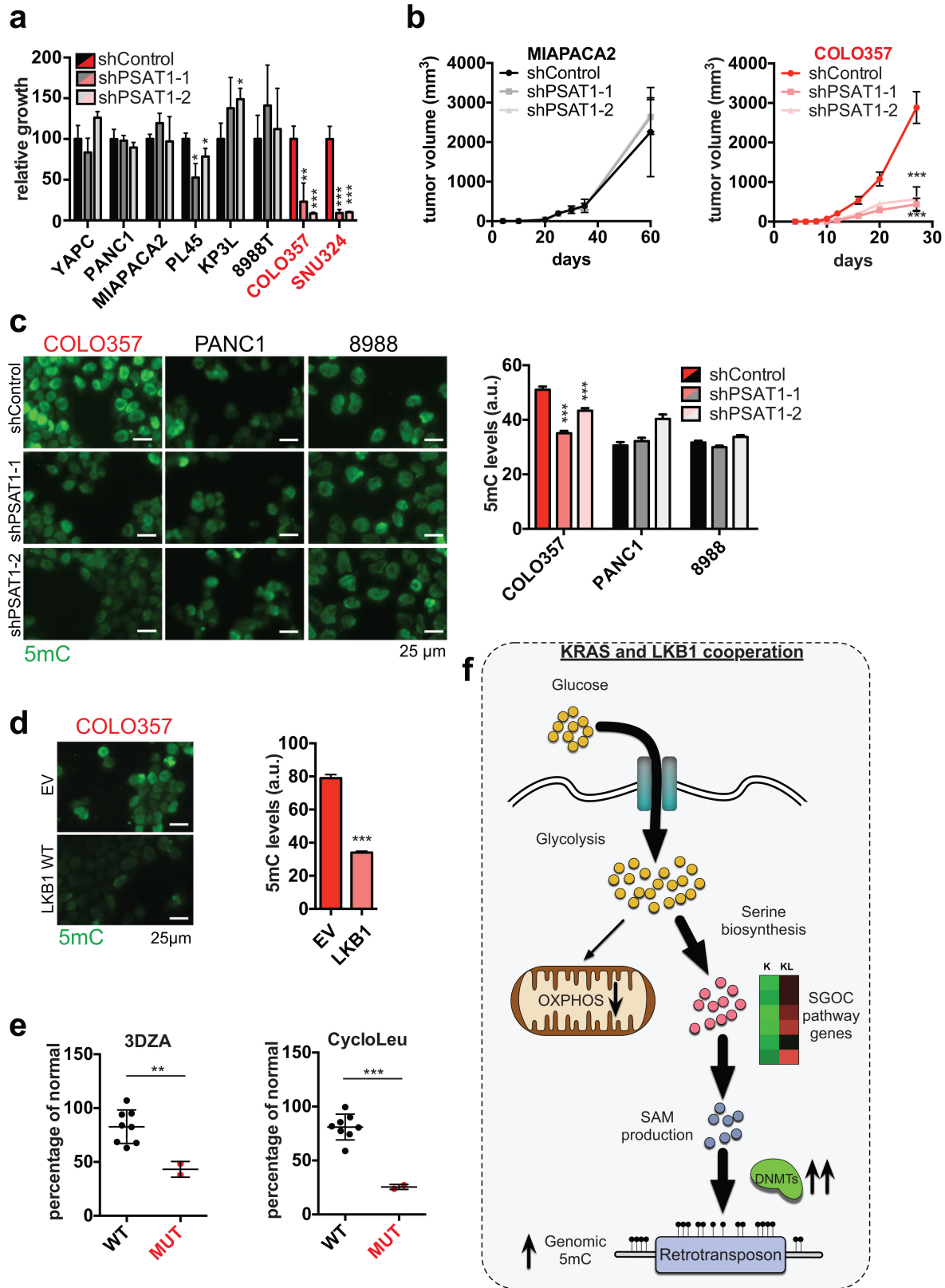
d–j, Distribution of methylation density within the low-CpG-promoters (LCPs), high-CpG-promoters (HCPs), islands, shores, exons, introns and LTRs. Numbers reflect median values. **k**, Average percentile change in methylation in the same elements as in **d–j** as well as LINEs and LTRs. **l**, Number of genes containing retrotransposon repeat elements (sum of LINEs, SINEs, LTRs) among the set of differentially regulated genes (DRGs) in K versus KL cells and among all genes. Note that the DRGs are enriched for the presence of retrotransposons in their gene bodies (top), but not in their promoters (middle) or 3'UTRs (bottom). **m**, Specific enrichment of LINE, SINE and LTR elements in the gene bodies of DRGs when compared to all genes in the genome.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | LKB1 deficiency confers hypersensitivity to inhibitors of DNA methylation. **a–c**, Proliferation of K (**a**), KL (**b**) and KPC (**c**) cells transduced with shControl or two shRNAs against each of DNMT1 (D1) or DNMT3A (D3A). Data are expressed as relative to day 0 (K and KL $n = 6$, KPC $n = 4$). **d–f**, Volume of subcutaneous tumours derived from KPC cells transduced with doxycycline (Dox)-inducible shRNAs against DNMT1 or DNMT3A ($n = 4$). Doxycycline was introduced to the drinking water when tumours reached 125 mm^3 . Error bars show s.e.m. For source data on tumour volume, see Supplementary Data Table 4. **g**, Apoptosis measured by caspase 3/7 activity in K or KL cells treated with decitabine for 48 h. Values are normalized to cell number ($n = 3$). **h, i**, Proliferation of KL cells transduced with empty vector (**h**) or LKB1 cDNA (**i**) treated with decitabine. Data are expressed as relative to day 0 ($n = 3$). **j**, Proliferation of KL cells transduced with empty vector or LKB1 cDNA treated with RG108. Data are expressed as percentage of growth of untreated cells ($n = 6$). **k–m**, Proliferation of K or KL cells treated with

RG108 ($n = 6$) (**k**), EGCG ($n = 12$) (**l**) or SGI1027 ($n = 12$) (**m**). Data are expressed as percentage of growth of untreated cells. **n–q**, Mice bearing subcutaneous KPC tumours were treated with decitabine ($n = 12$) or vehicle ($n = 12$) when tumours reached 125 mm^3 . **n, o**, Tumour volume (**n**) and final tumour weight (**o**). Error bars show s.e.m. For source data on tumour volume, see Supplementary Data Table 4. **p**, Haematoxylin and eosin-stained slides from representative tumours (top). Bottom, anti-CK19 (green) was used to visualize the neoplastic epithelium and anti-PCNA (red) was used to mark proliferating cells. DAPI was used to stain nuclei (blue). **q**, Quantification of the CK19⁺ neoplastic epithelial compartment (%CK19⁺ cells/total cells; top). Quantification of CK19⁺ cells with nuclear PCNA staining (bottom; $n = 6$). Scale bars, $100 \mu\text{m}$. Insets are threefold magnification. Data pooled from two (**a–c, j**) or four (**k–m**) experiments or representative of two (**h**) or three (**g**) experiments. Error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Vulnerabilities of human LKB1 mutant pancreatic cancer cell lines. **a**, Three-day growth of LKB1 wild-type (black) or LKB1 mutant (red) human pancreatic cancer cells. Data are expressed as relative to shControl-transduced cells, which is arbitrarily set to 100 ($n = 3$). **b**, Quantification of tumour volume of subcutaneous tumours derived from implantation of the indicated cells transduced with shControl or two shRNAs against PSAT1 ($n = 4$). Error bars show s.e.m. For source data on tumour volume, see Supplementary Data Table 4. **c**, Immunofluorescence staining and quantification of 5mC in COLO357, PANC1 and PATU-8988T cells transduced with shControl or two shRNAs against PSAT1 (607–760 cells for COLO357, 305–342 cells for PANC1 and 623–889 cells for PATU-8988T). Scale bar, 25 μ m. Error bars show s.e.m.

Data are expressed as fluorescence per nucleus. **d**, Detection of 5mC by immunofluorescence in COLO357 (LKB1-deficient) cells transduced with vector or wild-type LKB1. Quantification is presented as fluorescence per nucleus (438–618 cells). Scale bar, 25 μ m. Error bars show s.e.m. **e**, Three-day growth of LKB1 wild-type (black) or LKB1 mutant (red) human pancreatic cancer cells treated with 10 μ M 3-deazaadenosine (3DZA) (top) or 2 mM cycloleucine (CycloLeu) (bottom). **f**, Metabolic and epigenetic changes promoting transformation upon deletion of the tumour suppressor LKB1. Data pooled from (c, d) or representative of (a) two experiments. Error bars show s.d. unless otherwise stated; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

ARTICLE

doi:10.1038/nature20152

Thermophilic archaea activate butane via alkyl-coenzyme M formation

Rafael Laso-Pérez, Gunter Wegener, Katrin Knittel, Friedrich Widdel, Katie J. Harding, Viola Krukenberg, Dimitri V. Meier, Michael Richter, Halina E. Tegetmeyer, Dietmar Riedel, Hans-Hermann Richnow, Lorenz Adrian, Thorsten Reemtsma, Oliver Lechtenfeld & Florin Musat

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. *Nature* is providing this early version of the typeset paper as a service to our customers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Cite this article as: Laso-Pérez, R. *et al.* Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* <http://dx.doi.org/10.1038/nature20152> (2016).

Received 6 June; accepted 11 October 2016.

Accelerated Article Preview Published online 17 October 2016.

Thermophilic archaea activate butane via alkyl-coenzyme M formation

Rafael Laso-Pérez^{1,2}, Gunter Wegener^{1,2,3}, Katrin Knittel¹, Friedrich Widdel¹, Katie J. Harding^{1§}, Viola Krukenberg^{1,2}, Dimitri V. Meier¹, Michael Richter¹, Halina E. Tegetmeyer^{2,4}, Dietmar Riedel⁵, Hans-Hermann Richnow⁶, Lorenz Adrian⁶, Thorsten Reemtsma⁶, Oliver Lechtenfeld⁶ & Florin Musat^{1,6}

The anaerobic formation and oxidation of methane involve unique enzymatic mechanisms and cofactors that are believed to be all specific for C₁-compounds. Here we found that an anaerobic thermophilic enrichment culture composed of dense consortia of archaea and bacteria apparently uses partly similar pathways to oxidize the C₄-hydrocarbon butane. The archaea, proposed genus *Candidatus Syntrophoarchaeum*, showed the characteristic autofluorescence of methanogens, and contained highly expressed genes encoding enzymes similar to methyl-coenzyme M reductase (MCR). We detected butyl-coenzyme M, indicating archaeal butane activation in analogy to the first step in anaerobic methane oxidation. In addition, *Ca. Syntrophoarchaeum* expressed the genes encoding beta-oxidation enzymes, carbon monoxide dehydrogenase and reversible C₁-methanogenesis enzymes. This allows complete oxidation of butane. Reducing equivalents are apparently channelled to HotSeep-1, a thermophilic sulfate-reducing partner bacterium known from the anaerobic oxidation of methane. Genes encoding 16S rRNA and MCR similar to those identifying *Ca. Syntrophoarchaeum* were repeatedly retrieved from marine subsurface sediments suggesting that the presented activation mechanism is naturally widespread in the anaerobic oxidation of short-chain hydrocarbons.

Etymology. *Syntrophoarchaeum*. *syn* (greek): together, *tropho* (greek) meaning nourishment, *archaeum* (greek): ancient; butanivorans, *butanum* (latin): butane, *vorans* (latin): eating, devouring. The name implies an organism capable of butane oxidation, however demanding syntrophic electron sinks. A second strain is named *Ca. S. caldarius*, *caldarius* (latin): warm/ hot, refers to its thermophilic growth condition. **Locality.** Enriched from hydrothermally-heated, hydrocarbon-rich marine sediments of the Guaymas Basin at 2000 m water depth, Gulf of California, Mexico.

Diagnosis. Anaerobic, butane-oxidizing archaeon, variable morphology, 1.5 × 1 µm, dependent on syntrophic support by the sulfate-reducing partner bacterium *Candidatus Desulfoferriavidus auxilii*.

Massive amounts of natural gas migrate from deep-seated reservoirs towards the seafloor^{1–3}. Most of this gas is already consumed in the anoxic zone by microorganisms coupling oxidation of the hydrocarbons to the reduction of the abundant electron acceptor, sulfate. Research on the anaerobic oxidation of natural gas has been largely focused on methane as the most abundant constituent^{2–4}. The anaerobic oxidation of methane (AOM) is carried out by anaerobic methanotrophic archaea (ANME) forming consortia with sulfate-reducing bacteria (SRB); ANME activate methane to methyl-coenzyme M by methyl-CoM reductases (MCR)^{4,5}. The CoM-bound methyl group is further oxidized to CO₂ by reversing the enzymatic chain of methanogenesis^{6,7}, while reducing equivalents are channelled to the partner SRB^{8,9}.

However, natural gas generated by thermogenic decomposition of organic matter contains considerable amounts (up to 20%) short-chain alkanes including mostly ethane (up to 15%), but also propane, *iso*-butane and *n*-butane (hereafter butane)^{10,11}. Although these gases are potential growth substrates for microorganisms, the anaerobic

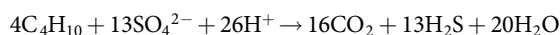
oxidation of the non-methane hydrocarbons became of interest only recently. All cultures known so far to oxidize short-chain hydrocarbons anaerobically are bacteria that couple complete substrate oxidation to CO₂ and sulfate reduction in one organism^{12–16}. The only pure culture available so far is strain BuS5, a deltaproteobacterium¹². Strain BuS5 and related organisms activate propane and butane via addition to fumarate yielding (1-methylalkyl)succinate, to date the most-studied anaerobic activation mechanism for hydrocarbons^{12,17–19}. A recent study, however, suggests that this mechanism is not always involved in short-chain hydrocarbon oxidation²⁰. Here we combined physiological experiments, community sequencing, microscopy, omics approaches and metabolite analyses to provide evidence for an alternative anaerobic activation reaction of butane in a highly enriched thermophilic culture of archaeal-bacterial consortia.

Archaea catalyse thermophilic butane oxidation

Thermophilic anaerobic methanotrophic archaea (ANME-1) were previously found to be associated with a deltaproteobacterial phylotype, HotSeep-1^{9,21} that had been detected independently in thermophilic enrichment cultures with short-chain alkanes^{12,15}. This prompted us to attempt again thermophilic enrichments with a short-chain hydrocarbon, butane; for these, we used inocula from Guaymas Basin sediment samples that had already been incubated with methane and developed thermophilic AOM activity. Indeed, after two months, also butane-dependent sulfate reduction became obvious. Further cultivation and subsequent transfers to new medium with butane yielded a sediment-free culture, hereafter called Butane50. This culture reduced 10 mM sulfate within approximately 14 weeks of incubation. Quantitative growth experiments showed that the amount of butane degraded was apparently coupled to the stoichiometric reduction of

¹Max-Planck Institute for Marine Microbiology, 28359 Bremen, Germany. ²Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, 27570 Bremerhaven, Germany. ³MARUM, Center for Marine Environmental Sciences, University Bremen, 28359 Bremen, Germany. ⁴Center for Biotechnology, Bielefeld University, 33615 Bielefeld, Germany. ⁵Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. ⁶Helmholtz Centre for Environmental Research – UFZ, 04318 Leipzig, Germany. [§]current address: University of California at Santa Cruz, Ocean Sciences Department, Santa Cruz, CA 95064, USA.

sulfate to sulfide in a 1:3.25 ratio, as demonstrated also for strain BuS5 (Figure 1a; Supplementary Table 1), yielding the net reaction



In this culture the ANME-1 phylotype was no longer detectable by molecular analysis. Instead, archaea closely related to *Methanosarcinales* had become dominant due to the change to butane as substrate; they were members of the GoM-Arch87 clade²² (Extended Data Table 1). Like ANME-1, also GoM-Arch87 forms densely packed consortia with the HotSeep-1 partner bacterium (Figure 1b,c). These consortia show strong blue-green autofluorescence (maximum at approx. 475 nm; Figure 1d), which is characteristic of cofactor F_{420} , a hydrogen carrier typical for methanogenic and ANME archaea^{6,23}. Phylotype GoM-Arch87 was originally detected in cold-seep areas of the Gulf of Mexico²², and since then was also found in other marine environments, especially seep and vent areas (see SILVA database release SSU 119; www.arb-silva.de). The environmental role of these archaea has been so far unknown.

From the metagenome assembly we retrieved two bins of contigs indicating genomes of two different GoM-Arch87 (Supplementary Table 2 and Extended Data Table 2). The two GoM-Arch87 bins have sizes of 1.46 Mbps (GoM-Arch87-1) and 1.66 Mbps (GoM-Arch87-2). Based on their tRNA content, archaeal specific single copy gene numbers and lineage-specific marker gene sets (*Euryarchaeota*) the draft genomes cover an estimated 85–89% of the complete genome of GoM-Arch87-1 and 95–97% of GoM-Arch87-2, indicating genome sizes of 1.64 to 1.71 Mbps. According to their 16S rRNA gene identity of 96% and whole genome identity of 74–90% (Figure 1e; Extended Data Table 2), the two bins likely represent different species of one genus. Hence we propose to name the more abundant (according to genome coverage) strain GoM-Arch87-1, *Candidatus Syntrophoarchaeum butanivorans* and the less abundant strain GoM-Arch87-2, *Candidatus Syntrophoarchaeum caldarius*. The two strains have largely similar gene content, and are further discussed together as *Ca. Syntrophoarchaeum*.

Ca. Syntrophoarchaeum encodes a fatty acid oxidation pathway, two complete acetyl-CoA decarboxylase/synthase:CO dehydrogenase (ACDS/CODH) complexes, and an almost complete methanogenesis-related pathway. Each of the two draft genomes contains four complete *mcr* gene sets, of which three (*Ca. S. butanivorans*) or all four (*Ca. S. caldarius*) are organized in operons (Extended Data Figure 1). The correct assembly and the unusually high number of *mcr* genes in a single genome were confirmed by cloning and resequencing each of the *mcrA* genes. Metatranscriptome and metaproteome analyses (only studied on *Ca. S. butanivorans*) showed that all these gene/protein sets were highly expressed/abundant (Extended Data Tables 3–5). Neither the metagenomic assembly nor the draft genomes of *Ca. Syntrophoarchaeum* contain genes encoding the otherwise wide-spread glycol radical enzymes (*assA/masD*, *bssA*) for anaerobic hydrocarbon activation. Accordingly, these genes could not be amplified from extracted DNA using specific primers (see methods). These results strongly indicated that the activation of butane in *Ca. Syntrophoarchaeum* proceeds via mechanisms different than those described for bacteria. An appealing hypothesis was that butane in *Ca. Syntrophoarchaeum* is activated by the distinct variants of MCR in a reaction analogous to that of methane at ‘reverse’ MCR in ANME. The expected first metabolite would thus be butyl-coenzyme M rather than methyl-coenzyme M.

Ca. Syntrophoarchaeum forms butyl-coenzyme M

Analysis of cell extracts of active Butane50 cultures by direct infusion ultra-high resolution mass spectrometry indeed provided direct evidence for butane-dependent formation of butyl-coenzyme M. A mass peak of $m/z = 197.03116$ was detected, which corresponded exactly to the hypothesized butyl-CoM ($\text{C}_6\text{H}_{13}\text{S}_2\text{O}_3^-$). Following quadrupole isolation and collision-induced fragmentation this compound yielded two major fragments representing a butylthiol ($m/z = 89.0430$, $\text{C}_4\text{H}_9\text{S}^-$) and an apparently CoM-derived bisulfite ($m/z = 80.9652$, HSO_3^- ; Figure 2

and Extended Data Figure 2a). Use of synthesized authentic standards showed that the two isomers, 1-butyl-CoM and 2-butyl-CoM, were present and separated by liquid chromatography, with apparently higher amounts of the latter (Figure 2c). The presence of two butyl-CoM isomers offers, in principle, two explanations. First, one could be the initial product whereas the other is the subsequently formed intermediate. Second, only one isomer may be the genuine, directly metabolized activation product, whereas the other is formed as a by-product due to relaxed catalytic specificity of the activating enzyme; this would be similar to the anaerobic activation of propane by strain BuS5, where both *n*- and *iso*-propylsuccinate have been detected as activation products^{12,24}. Because the different amounts of the detected isomers reflect steady state concentrations (dynamic pools), they cannot offer clues to their metabolic significance. To further corroborate butyl-CoM formation as a specific activation reaction, i.e. exclude that it can be accomplished ‘accidentally’ by other means, we added butane to a methane-oxidizing ANME-1 enrichment. Neither butyl-CoM nor its derived fragments were detected. Moreover, these compounds were not detected in sterile controls, in substrate-starved Butane50 cultures, or in cultures of the bacterial butane-oxidizer strain BuS5 (Figure 2 and Extended Data Figure 2a).

Mechanistic relatedness between butane activation at the MCR variant and methane activation at MCR of methanotrophic archaea was also shown by an inhibition experiment. The coenzyme M analogue, bromoethanesulfonate (BES), which inhibits methanogenesis as well as anaerobic methane oxidation²⁵ also inhibited butane-dependent sulfate reduction (Extended Data Figure 3). Inhibition must occur in the archaeal metabolism, because BES did not affect the partner bacterium, HotSeep-1 (*Ca. Desulfofervidus auxilii*) when separately grown with hydrogen and sulfate; the tested partner originated from the thermophilic AOM culture²⁶.

Aliquots of the Butane50 culture were also tested with other hydrocarbons. After approx. two months of incubation with propane, sulfate reduction became obvious. The rate increased and became similar as with butane. A possible shift in the microbial composition or gene expression upon incubation with propane has yet not been assessed. However, a mass peak exactly corresponding to that of propyl-CoM was analysed ($m/z = 183.0155$; Extended Data Figure 2c). This further corroborates the formation of alkyl-CoM as the apparent initial product in short-chain alkane activation. So far, no sulfate reduction was detectable with added methane, ethane, *iso*-butane, *n*-pentane or *n*-hexane.

This is the first demonstration of coenzyme M acting *in vivo* as a carrier of alkyl moieties other than the methyl group. The closest known precedent is the *in vitro* reduction of the homologous substrate ethyl-CoM to ethane by MCR from methanogenic archaea^{27,28}. Turnover rates of ethyl-CoM in these assays were two orders lower than of methyl-CoM, which was explained by a steric hindrance of the larger ethyl-CoM molecule in the MCR enzyme²⁸. This suggests that an efficient activation of butane to butyl-CoM may require highly adapted MCR-like enzymes. Analysis of the deduced amino acid sequences of the eight *mcr* genes detected in *Ca. Syntrophoarchaeum* showed a low similarity among them with an identity of only 29–82%. Two different MCR enzymes encoded in the genome of *Ca. S. butanivorans* were highly abundant in protein extracts of the culture, qualifying them as likely candidates for butane activation (Extended Data Table 3).

A high substrate specificity of the MCR type apparently involved in butane oxidation was also indicated in an incubation experiment with both, butane and methane. In the active culture, butyl-CoM but no methyl-CoM was detected (Extended Data Figure 2b). Hence, there was no evidence for co-activation of methane, even though alone by its molecular size the latter should not encounter any steric binding hindrance. One may speculate that reaction at the enzyme requires substrate ‘fixation’ by binding of an alkyl chain of a minimum length. Understanding as to which extent the slightly weaker C–H-bond in higher alkanes in comparison to that in methane could play a role may need refined theoretical consideration. Furthermore, the population

and abundant MCR type in the recent propane-adapted subculture will have to be analysed. At the present state it is unknown whether propane and butane are activated by the same organism and enzyme or cause selection of different ones with respective substrate preference.

Notably three of the McrA subunits of *Ca. S. butanivorans* have a related equivalent in *Ca. S. caldarius*. Interestingly the fourth McrA subunit (SBU_000314) forms a cluster with McrA sequences of the recently described *Bathyarchaeota*²⁹, whereas the fourth sequence of *Ca. S. caldarius* (SCAL_000352) has a distant phylogenetic position (Figure 3; for further phylogenetic analysis see Methods and Supplementary Figure 1). Together McrA sequences of *Ca. Syntrophoarchaeum* and *Bathyarchaeota* form a highly divergent cluster compared to those of methanogens or methanotrophs, which is likely due to the ability of the encoded enzymes to efficiently accommodate larger substrates such as butane.

Complete oxidation of butane

The measured stoichiometric balance shows that the CoM-bound butyl moiety must be completely degraded to CO₂. Only a minor fraction is expected to enter biosynthesis, like in other strict anaerobes³⁰. Complete oxidation requires four basic metabolic features, the conversion of the butyl-thioester to the butyryl-thioester (presumably butyryl-CoA), oxidation of butyryl-CoA to acetyl-CoA, terminal oxidation of acetyl-CoA to CO₂, and channelling of the reducing equivalents (electrons) into sulfate reduction. Our proteogenomic analysis provides explanations for all of these processes, with the exception of the first.

The role and significance of the butyl-CoM isomers and their subsequent processing steps could not be clarified on the basis of our present analyses. Also, there are no precedent cases of natural pathways involving non-methyl CoM-thioesters that could suggest particular reactions. Still, one may argue that 1-butyl-CoM is the genuine and directly metabolized activation product. Because MCR in AOM activates a primary C—H-bond, its variant in culture Butane50 may also attack butane at the primary carbon. In this way, first oxidation and functionalization would occur at the same carbon atom which in a subsequent beta-oxidation would become the activated carboxyl group (thioester). In such case, 2-butyl-CoM would be a by-product; it is presently unknown whether an isomerase converts it to 1-butyl-CoM.

In methyl-CoM oxidation in AOM and in the oxidative branch in methanol-grown *Methanosarcina*, the subsequent steps would be transfer of the methyl group to tetrahydromethanopterin (H₄MPT) or tetrahydrosarcinapterin (H₄SPT) and oxidation via the bound formaldehyde and formate states finally to CO₂. None of the genes encoding methyl-H₄MPT:coenzyme M methyltransferase were found in the genome. An analogous transfer of the butyl-moiety by the abundant, presumably modified methyltransferase (Mta) may be only speculated about (Figure 4, Extended Data Table 3), because the postulated butyl-cobalamin intermediate would be unprecedented. Furthermore, tetrahydropterins and methanofuran are, to our knowledge, exclusive C₁-carriers. Hence, the enzymatic reactions converting 1-butyl-CoM to the butyryl (bound acid) level are presently puzzling, even though transfer and oxidation reactions of higher carbon compounds can be theoretically formulated also with the C₁-carriers.

The predicted beta oxidation enzymes of *Ca. Syntrophoarchaeum* (Figure 4, Extended Data Table 3) are highly related to those of deltaproteobacterial sulfate reducers and syntrophic partners of methanogens^{31,32} (Extended Data Table 6). This suggests horizontal gene transfer across the two domains of life. The gene encoding the most abundantly formed acyl-CoA dehydrogenase of *Ca. Syntrophoarchaeum* is located in an operon with genes for an electron transfer flavoprotein (*etf*) complex and a [FeS]-oxidoreductase. Hence, these gene products likely act as electron acceptors in the oxidation of butyryl-CoA (Figure 4) as previously shown for bacterial butyrate oxidizers^{33,34}.

The key enzyme for the metabolism of acetyl-CoA from beta-oxidation is the detected acetyl-CoA decarboxylase/synthase:CO

dehydrogenase (ACDS/CODH). The carboxyl group-derived bound CO can be oxidized to free CO₂ and yield reduced ferredoxin as shown before for methanogens²³. The methyl group is probably oxidized via the reverse methanogenesis pathway, of which almost all genes are present and their corresponding transcripts and proteins were detected (Figure 4, Extended Data Table 4). The N5,N10-methylenetetrahydromethanopterin (methylene-H₄MPT) reductase (*mer*) gene is absent in *Ca. Syntrophoarchaeum*. However, like in ANME-1 Mer could be substituted by the highly abundant putative methylenetetrahydrofolate reductase (Met) complex as suggested before³⁵. The genes encoding the Met complex of *Ca. Syntrophoarchaeum* have a similar operon structure to those of *Moorella thermoacetica*³⁶. In conclusion, the combination of enzymes in *Ca. Syntrophoarchaeum* results in the archaeal version of the oxidative Wood-Ljungdahl pathway, as originally shown for the sulfate-reducing *Archaeoglobus fulgidus*³⁷.

Ca. Syntrophoarchaeum does not have the genes for canonical sulfate reduction (i.e. *dsrAB*, *aprAB*), and hence depends on an external electron sink. This role is apparently fulfilled by the partner bacterium. We tested for the production of hydrogen as a canonical intermediate in syntrophic associations. Although *Ca. Syntrophoarchaeum* encodes for a cytoplasmic [NiFe] hydrogenase (Extended Data Table 5), only minor amounts of hydrogen were produced in active or molybdate-treated Butane50 cultures, which cannot explain the required reducing equivalent transfer (Figure 5a). To further study how reducing equivalents are transferred to the partner bacterium, the draft genome of the Butane50 HotSeep-1 strain was retrieved from the metagenome. Based on different marker genes, this HotSeep-1 strain is basically identical to *Ca. D. auxilii*, the partner in thermophilic AOM²⁶ (Supplementary Table 3). *Ca. D. auxilii* is a lithoautotrophic sulfate reducer that in thermophilic AOM thrives on electrons directly supplied by the ANME via pili-based nanowires and cytochromes^{9,26}. Also in the Butane50 culture HotSeep-1 expresses genes encoding pili assembly proteins, especially the main component of type IV pilus (PilA), and many different potentially cytochromes including extracellular ones (Extended Data Table 7). Indeed, using transmission electron microscopy we found a similarly dense, apparently pili-based, nanowire network in the intercellular space of the Butane50 consortia (Figure 5b). Based on these results we propose that nanowire-based direct interspecies electron transfer³⁸ mediates also electron exchange in the butane-oxidizing consortia. From a thermodynamic point of view, the interspecies transfer of reducing equivalents would be also possible via H₂; its measured partial pressure of ca. 1 Pa would allow energy conservation in both partners, albeit with quite unequal energy sharing (calculation in Supplementary Discussion; Supplementary Figure 2). However, the extremely slow rate of H₂ formation (Figure 5a) is difficult to reconcile with the observed rate of sulfate reduction.

Channelling of the reducing equivalents to the sulfate-reducing partner bacterium requires electrons with redox potentials below the average potential of sulfate reduction ($E^{\circ} = -220$ mV), which is reached by most proposed redox reactions. However, the oxidation of butyryl-CoA to crotonyl-CoA (Figure 4), will release electrons with far higher potentials ($E^{\circ} = -125/-10$ mV³³). A shift in the redox potential of these electrons can be achieved by different mechanisms, including energy-driven reverse electron transport³³ or electron confurcation³⁹, both demonstrated for syntrophic bacteria. Two of the required genes, encoding EtfAB and FeS oxidoreductase, were identified in *Ca. Syntrophoarchaeum*. Other genes with products involved in electron transport have been identified including different heterodisulfide reductases (Extended Data Table 5) like one complex of HdrABC (SBUT_000297-299) followed by the genes encoding the β -subunit of formate dehydrogenase (FdhB) and the δ -subunit of methylviologen-reducing hydrogenase (MvhD) (Extended Data Table 5). This complex was already identified in ANME-1 genomes and was hypothesised to be involved as electron-accepting complex⁷.

Evolutionary and environmental aspects

The analysis of the anaerobic consortia enriched with butane provides first insights into an additional mechanism for the oxygen-independent activation and use of non-methane saturated hydrocarbons in archaea. Saturated hydrocarbons belong to the chemically least reactive compounds. In anaerobic bacteria, the radical-catalysed addition of alkanes to fumarate yielding substituted succinates is regarded as common^{40,41}, even though alternative mechanisms have been proposed^{42,43}. The MCR-like proteins produced by *Ca. Syntrophoarchaeum* and formation of butyl-coenzyme M in the culture suggest that the mechanistic principle for the activation of methane, the most stable hydrocarbon, was evolutionarily adapted to use also short chain alkanes. Another requirement for metabolic processing was, apart from the yet unsolved subsequent step of the alkyl-thioether, the acquisition of enzymes for beta-oxidation, which has been shown before in other archaea⁴⁴. The subsequent reactions leading to CO₂ are enabled by steps which in principle are long-known in methanogens; acetate- and methanol-utilizing species contain CO-dehydrogenase/acetyl-CoA synthase as well as C₁-processing enzymes able to operate in oxidative direction⁴⁵. Such an archaeal variant of the Wood-Ljungdahl pathway for acetyl-CoA oxidation was first described for the sulfate reducer *Archaeoglobus fulgidus*³⁷.

The presently suggested pathway for anaerobic oxidation of non-methane alkanes might not be limited to *Ca. Syntrophoarchaeum*. The recently detected uncultivated *Bathyarchaeota* contain genes for MCR types highly similar to those of *Ca. Syntrophoarchaeum* (Figure 3). Furthermore, the draft genomes of *Bathyarchaeota*²⁹ as well as the pan-genome of uncultivated *Hadesarchaea*⁴⁶ encode most enzymes of the methanogenesis pathway, for short-chain fatty acid oxidation and for acetyl-CoA oxidation via ACDS/CODH. Based on these predicted features, *Bathyarchaeota* were described as heterotrophic methanogens. Our results indicate that at least some of these uncultivated archaea may be also viewed as alkane oxidizers.

In conclusion, anaerobic microorganisms thriving on non-methane alkanes may be phylogenetically and metabolically even more diverse than hitherto thought^{47,48}. Such diversity may reflect a long-existing presence of saturated hydrocarbons in the biosphere and their exploitation as growth substrates through various evolving pathways, long before the oxygen-dependent ones. The factors which in modern anoxic environments select for consortial, archaeal hydrocarbon oxidation over an oxidation by single bacterial species^{12,41} are still unknown and need more comprehensive *in situ* studies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 June; accepted 11 October 2016.

Published online 17 October 2016.

- Simoneit, B. R. T., Kawa, O. E. & Brault, M. Origin of Gases and Condensates in the Guaymas Basin Hydrothermal System (Gulf of California). *Chemical geology* **71**, 169–182, doi:10.1016/0009-2541(88)90113-1 (1988).
- Reeburgh, W. Oceanic methane biogeochemistry. *Chem Rev.* **107**, 486–513 (2007).
- Boetius, A. & Wenzhöfer, F. Seafloor oxygen consumption fuelled by methane from cold seeps. *Nature Geoscience* **6**, 725–734, doi:10.1038/NGEO1926 (2013).
- Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**, 484–487, doi:10.1126/science.1061338 (2001).
- Boetius, A. et al. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623–626, doi:10.1038/35036572 (2000).
- Hallam, S. J. et al. Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457–1462, doi:10.1126/science.1100025 (2004).
- Meyerdierks, A. et al. Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ Microbiol* **12**, 422–439, doi:10.1111/j.1462-2920.2009.02083.x (2010).
- McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature advance online publication*, doi:10.1038/nature15512 (2015).

- Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E. & Boetius, A. Inter-cellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526**, 587–590 (2015).
- Claypool, G. E. & Kvenvolden, K. A. Methane and Other Hydrocarbon Gases in Marine Sediment. *Annu Rev Earth Planet Sc* **11**, 299–327, doi:DOI 10.1146/annurev.11.050183.001503 (1983).
- Tissot, B. P. & Welte, D. H. *Petroleum formation and occurrence*. (Springer-Verlag, 1984).
- Kniemeyer, O. et al. Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria. *Nature* **449**, 898–901, doi:10.1038/nature06200 (2007).
- Savage, K. N. et al. Biodegradation of low-molecular-weight alkanes under mesophilic, sulfate-reducing conditions: metabolic intermediates and community patterns. *FEMS Microbiol Ecol* **72**, 485–495 (2010).
- Jaekel, U. et al. Anaerobic degradation of propane and butane by sulfate-reducing bacteria enriched from marine hydrocarbon cold seeps. *ISME J* **7**, 885–895, doi:10.1038/ismej.2012.159 (2013).
- Adams, M. M., Hoarfrost, A. L., Bose, A., Joye, S. B. & Gurgis, P. R. Anaerobic oxidation of short-chain alkanes in hydrothermal sediments: potential influences on sulfur cycling and microbial diversity. *Front. Microbiol.* **4**, 110, doi:10.3389/fmicb.2013.00110 (2013).
- Kleindienst, S. et al. Diverse sulfate-reducing bacteria of the Desulfosarcina/Desulfococcus clade are the key alkane degraders at marine seeps. *ISME J* **8**, 2029–2044, doi:10.1038/ismej.2014.51 (2014).
- Kropp, K. G., Davidova, I. A. & Suflita, J. M. Anaerobic oxidation of n-dodecane by an addition reaction in a sulfate-reducing bacterial enrichment culture. *Applied and environmental microbiology* **66**, 5393–5398 (2000).
- Rabus, R. et al. Anaerobic initial reaction of n-alkanes in a denitrifying bacterium: evidence for (1-methylpentyl) succinate as initial product and for involvement of an organic radical in n-hexane metabolism. *Journal of Bacteriology* **183**, 1707–1715 (2001).
- Heider, J. et al. Structure and function of benzylsuccinate synthase and related fumarate-adding glycol radical enzymes. *Journal of Molecular Microbiology and Biotechnology* **26**, 29–44 (2016).
- Gittel, A. et al. Ubiquitous presence and novel diversity of anaerobic alkane degraders in cold marine sediments. *Frontiers in Microbiology* **6** (2015).
- Holler, T. et al. Thermophilic anaerobic oxidation of methane by marine microbial consortia. *ISME J* **5**, 1946–1956, doi:10.1038/ismej.2011.77 (2011).
- Orcutt, B. N. et al. Impact of natural oil and higher hydrocarbons on microbial diversity, distribution, and activity in Gulf of Mexico cold-seep sediments. *Deep-Sea Research Part II-Topical Studies in Oceanography* **57**, 2008–2021, doi:10.1016/j.dsr2.2010.05.014 (2010).
- Thauer, R. Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture. *Microbiology* **144**, 2377–2406 (1998).
- Jaekel, U., Vogt, C., Fischer, A., Richnow, H.-H. & Musat, F. Carbon and hydrogen stable isotope fractionation associated with the anaerobic degradation of propane and butane by marine sulfate-reducing bacteria. *Environmental Microbiology* **16**, 130–140, doi:10.1111/1462-2920.12251 (2014).
- Nauhaus, K., Treude, T., Boetius, A. & Kruger, M. Environmental regulation of the anaerobic oxidation of methane: a comparison of ANME-I and ANME-II communities. *Environ Microbiol* **7**, 98–106 (2005).
- Krukenberg, V. et al. Candidatus Desulfotomaculum auxilii, a hydrogenotrophic sulfate-reducing bacterium of the HotSeep-1 cluster involved in the thermophilic anaerobic oxidation of methane. *Environ Microbiol* (2016).
- Ahn, Y., Krzycki, J. A. & Floss, H. G. Steric course of the reduction of ethyl coenzyme M to ethane catalyzed by methyl coenzyme M reductase from *Methanosarcina barkeri*. *Journal of the American Chemical Society* **113**, 4700–4701 (1991).
- Scheller, S., Goenrich, M., Thauer, R. K. & Jaun, B. Methyl-coenzyme M reductase from methanogenic archaea: isotope effects on label exchange and ethane formation with the homologous substrate ethyl-coenzyme M. *J. Am. Chem. Soc* **135**, 14985–14995 (2013).
- Evans, P. N. et al. Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438, doi:10.1126/science.aac7745 (2015).
- Nauhaus, K., Albrecht, M., Elvert, M., Boetius, A. & Widdel, F. In vitro cell growth of marine archaeal-bacterial consortia during anaerobic oxidation of methane with sulfate. *Environ Microbiol* **9**, 187–196 (2007).
- McInerney, M., Bryant, M., Hespell, R. & Costerton, J. *Syntrophomonas wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium. *Applied and environmental microbiology* **41**, 1029–1039 (1981).
- Schmidt, J. E. & Ahning, B. K. Effects of hydrogen and formate on the degradation of propionate and butyrate in thermophilic granules from an upflow anaerobic sludge blanket reactor. *Applied and environmental microbiology* **59**, 2546–2551 (1993).
- Schmidt, A., Müller, N., Schink, B. & Schleheck, D. A proteomic view at the biochemistry of syntrophic butyrate oxidation in *Syntrophomonas wolfei*. *PLoS one* **8**, e56905 (2013).
- Wasmund, K. et al. Genome sequencing of a single cell of the widely distributed marine subsurface Dehalococcoidia, phylum Chloroflexi. *ISME J* **8**, 383–397, doi:10.1038/ismej.2013.143 (2014).
- Stokke, R., Roalkvam, I., Lanzén, A., Hafliðason, H. & Steen, I. H. Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. *Environ Microbiol* **14**, 1333–1346, doi:10.1111/j.1462-2920.2012.02716.x (2012).

36. Mock, J., Wang, S., Huang, H., Kahnt, J. & Thauer, R. K. Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in *Moorella thermoacetica*. *Journal of Bacteriology* **196**, 3303–3314 (2014).
37. Möller-Zinkhan, D. & Thauer, R. K. Anaerobic lactate oxidation to 3 CO₂ by *Archaeoglobus fulgidus* via the carbon monoxide dehydrogenase pathway: demonstration of the acetyl-CoA carbon-carbon cleavage reaction in cell extracts. *Archives of microbiology* **153**, 215–218 (1990).
38. Summers, Z. M. *et al.* Direct Exchange of Electrons Within Aggregates of an Evolved Syntrophic Coculture of Anaerobic Bacteria. *Science* **330**, 1413–1415, doi:10.1126/science.1196526 (2010).
39. Weghoff, M. C., Bertsch, J. & Müller, V. A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environmental Microbiology* **17**, 670–677 (2015).
40. Callaghan, A. V. Metabolomic investigations of anaerobic hydrocarbon-impacted environments. *Curr Opin Biotech* **24**, 506–515, doi:10.1016/j.copbio.2012.08.012 (2013).
41. Musat, F. The anaerobic degradation of gaseous, nonmethane alkanes—From in situ processes to microorganisms. *Computational and structural biotechnology journal* **13**, 222–228 (2015).
42. So, C. M., Phelps, C. D. & Young, L. Anaerobic transformation of alkanes to fatty acids by a sulfate-reducing bacterium, strain Hxd3. *Appl. Environ. Microbiol.* **69**, 3892–3900 (2003).
43. Heider, J., Szaleniec, M., Suenwoldt, K. & Boll, M. Ethylbenzene dehydrogenase and related molybdenum enzymes involved in oxygen-independent alkyl chain hydroxylation. *Journal of Molecular Microbiology and Biotechnology* **26**, 45–62 (2016).
44. Klenk, H.-P. *et al.* The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370 (1997).
45. Ferry, J. G. Biochemistry of methanogenesis. *Critical Reviews in Biochemistry and Molecular Biology* **27**, 473–503 (1992).
46. Baker, B. J. *et al.* Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nature Microbiology*, 16002, doi:10.1038/nmicrobiol.2016.2(2016).
47. Stagars, M. H., Ruff, S. E., Amann, R. & Knittel, K. High diversity of anaerobic alkane-degrading microbial communities in marine seep sediments based on (1-methylalkyl) succinate synthase genes. *Frontiers in Microbiology* **6** (2015).
48. Rabus, R., Boll, M., Golding, B. & Wilkes, H. Anaerobic Degradation of p-Alkylated Benzoates and Toluenes. *Journal of Molecular Microbiology and Biotechnology* **26**, 63–75 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Ramona Appel, Kathrin Büttner, Ines Kattelmann and Susanne Menger for assistance in cultivation and molecular work, Benjamin Scheer for proteomic analyses, and Marcel Sickert, Jakob Then Bergh and Kristina Dürkop for support in standard synthesis and LC-MS/MS analyses. We thank as well Antonio Fernández-Guerra, Harald Gruber-Vodicka and Pierre Offre for supporting us in bioinformatics and biochemistry. We thank Antje Boetius for fruitful discussions and financial support by her Leibniz Grant of the Deutsche Forschungsgemeinschaft (DFG). Research was further financed by the DFG Research Center and Cluster of Excellence MARUM and the Deep Carbon Observatory (Deep Life grant 11121/6152-2121-2329-9973-CC to G.W.), the Max Planck Society and the Helmholtz Society. We are indebted to Andreas Teske and the shipboard party, the crew and pilots of research expedition AT15-16 Research Vessel Atlantis and Research Submersible Alvin (NSF Grant OCE-0647633). We acknowledge the Centre for Chemical Microscopy (ProVIS) at the Helmholtz Centre for Environmental Research supported by European regional Development Funds (EFRE – Europe funds Saxony) for using their analytical facilities.

Author Contributions G.W. and F.M. retrieved the original samples and performed cultivation. R.L.-P., G.W., and F.M. designed research. R.L.-P., K.K., K.J.H. and V.K. designed the CARD-FISH probes and performed microscopy. R.L.-P., G.W. and F.M. performed physiological experiments. H.E.T. prepared and sequenced the DNA and RNA libraries. R.L.-P., V.K., D.V.M. and M.R. performed metagenomic and transcriptomic analyses. R.L.-P., K.K. and K.J.H. performed phylogenetic analysis. D.R. performed thin-sectioning and electron microscopy. H.-H.R., L.A. and F.M. performed proteome analyses. T.R., O.L. and F.M. analysed metabolic intermediates. R.L.-P., G.W., F.W. and F.M. developed the metabolic model, and wrote the manuscript with contributions of all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W. (gwegener@mpi-bremen.de) or F.M. (florin.musat@ufz.de)

Reviewer Information *Nature* thanks T. Ettema, S. Ragsdale, R. Thauer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

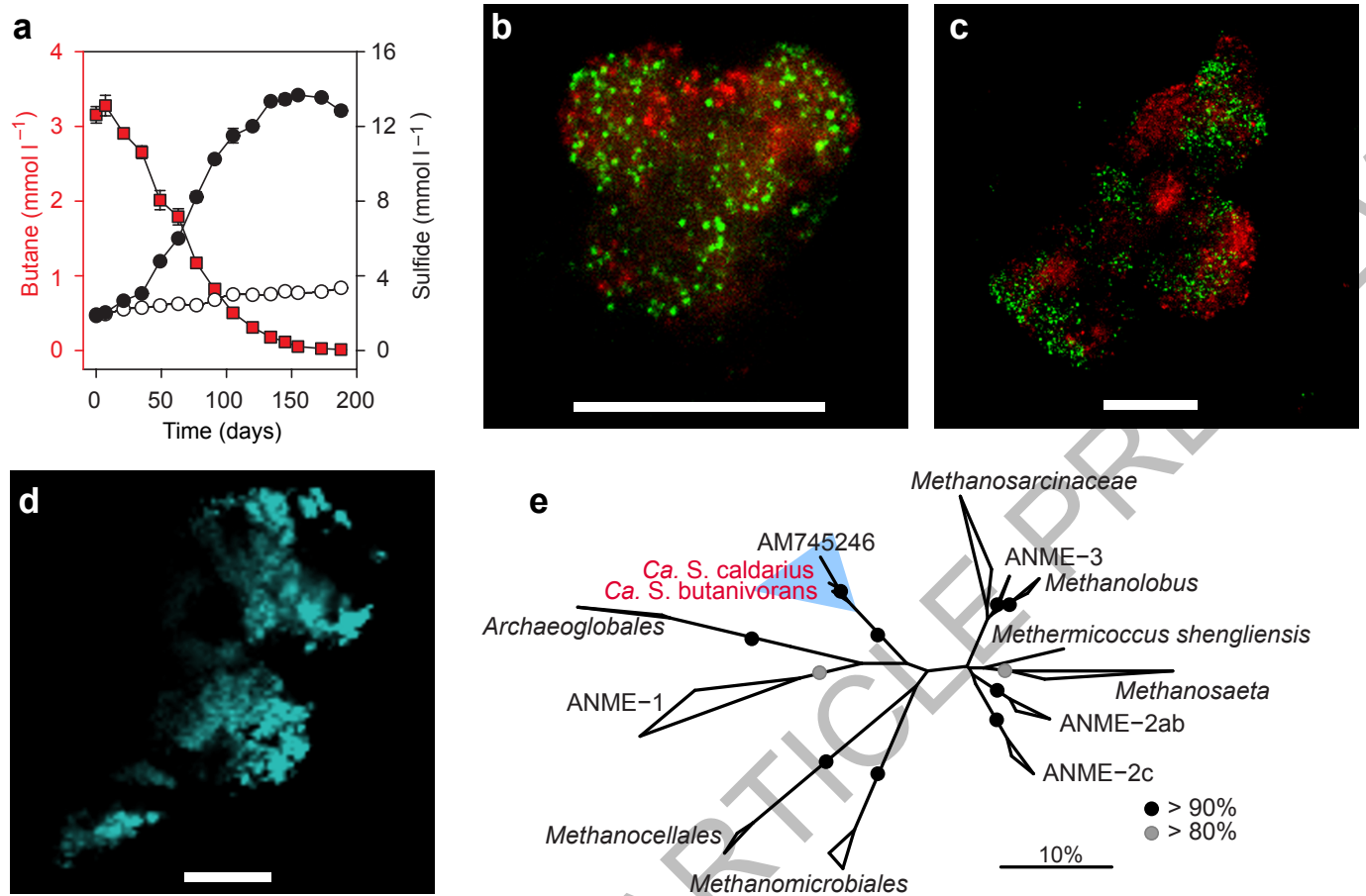


Figure 1 | Characterization of the Butane50 culture. **a**, The Butane50 culture consumed butane (red squares) coupled with stoichiometric reduction of sulfate to sulfide (black circles); no sulfide was produced in cultures without butane (white circles), error bar = standard deviation, $n = 3$. **b,c**, Fluorescence micrographs of Butane50 consortia stained with specific probes for GoM-Arch87 (red) and HotSeep-1 (green); representative for 20 recorded images, scale bar = 10 μm . **d**, Autofluorescence of microbial consortia visualized using excitation

light at 405 nm and a longpass emission filter (>463 nm). The autofluorescence maximum at 470 nm is indicative of the presence of the cofactor F_{420} ; representative for 10 recorded images, scale bar = 10 μm . **e**, Phylogenetic affiliation of the 16S rRNA gene sequences from the studied *Ca. Syntrophoarchaeum* strains (in red) within the *Euryarchaeota*. The GoM-Arch87 cluster is indicated with blue background; bar = 10% estimated sequence divergence; bootstraps values >80% and >90% are indicated by grey and black circles, respectively.

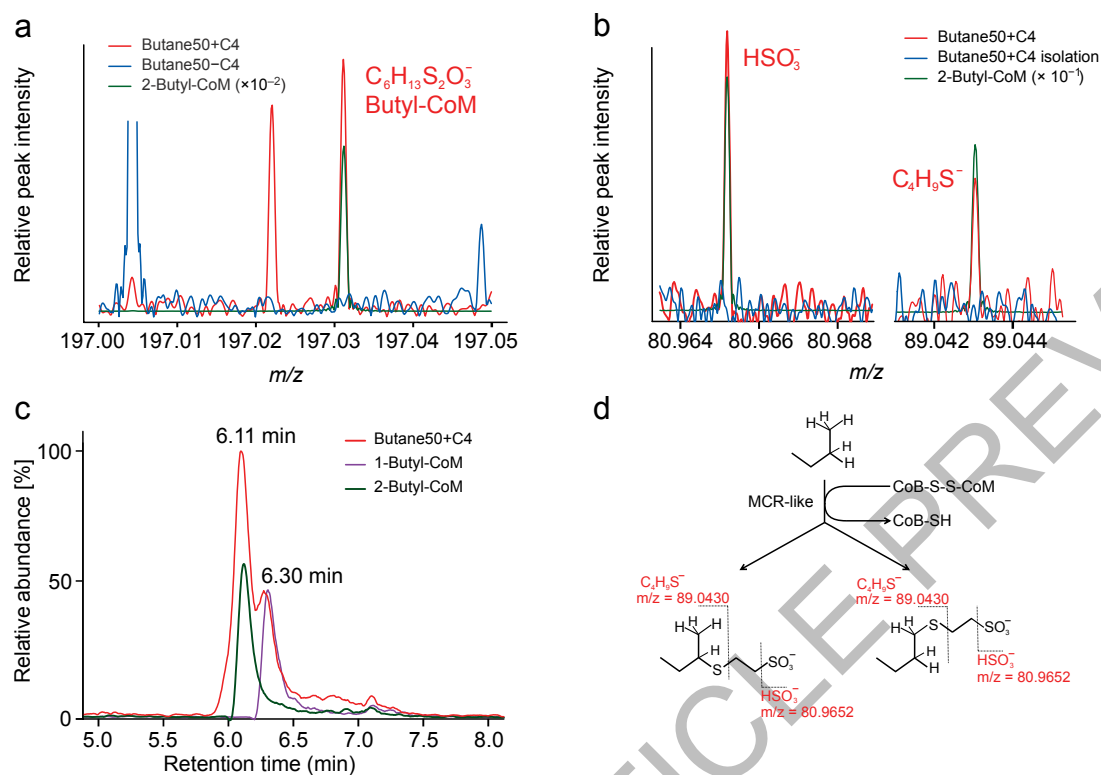


Figure 2 | Butyl-CoM as initial metabolic intermediate in butane oxidation **a**, Full scan mass spectrum of Butane50 culture extracts (Butane50+C4; $n = 6$ at different time points) revealed a peak at $m/z = 197.0312$ which matches the butyl-CoM standard (mass accuracy $+0.18$ ppm). This mass peak was absent in control incubations without butane (Butane50-C4; for more controls see Extended Data Figure 2). **b**, Isolation and collision induced fragmentation of this mass peak yielded

butylthiol (mass accuracy $+0.01$ ppm) and bisulfite (mass accuracy $+0.03$ ppm), which are likewise produced by fragmentation of the butyl-CoM standards. **c**, Liquid chromatography resolved the presence of the two isomers 1- and 2-butyl-CoM in the culture ($n = 4$ at different sulfide concentrations). **d**, Interpretation of these analyses. Butane is activated by ligation to coenzyme M by MCR-like enzymes, yielding both 1- and 2-butyl-CoM with identical fragmentation patterns.

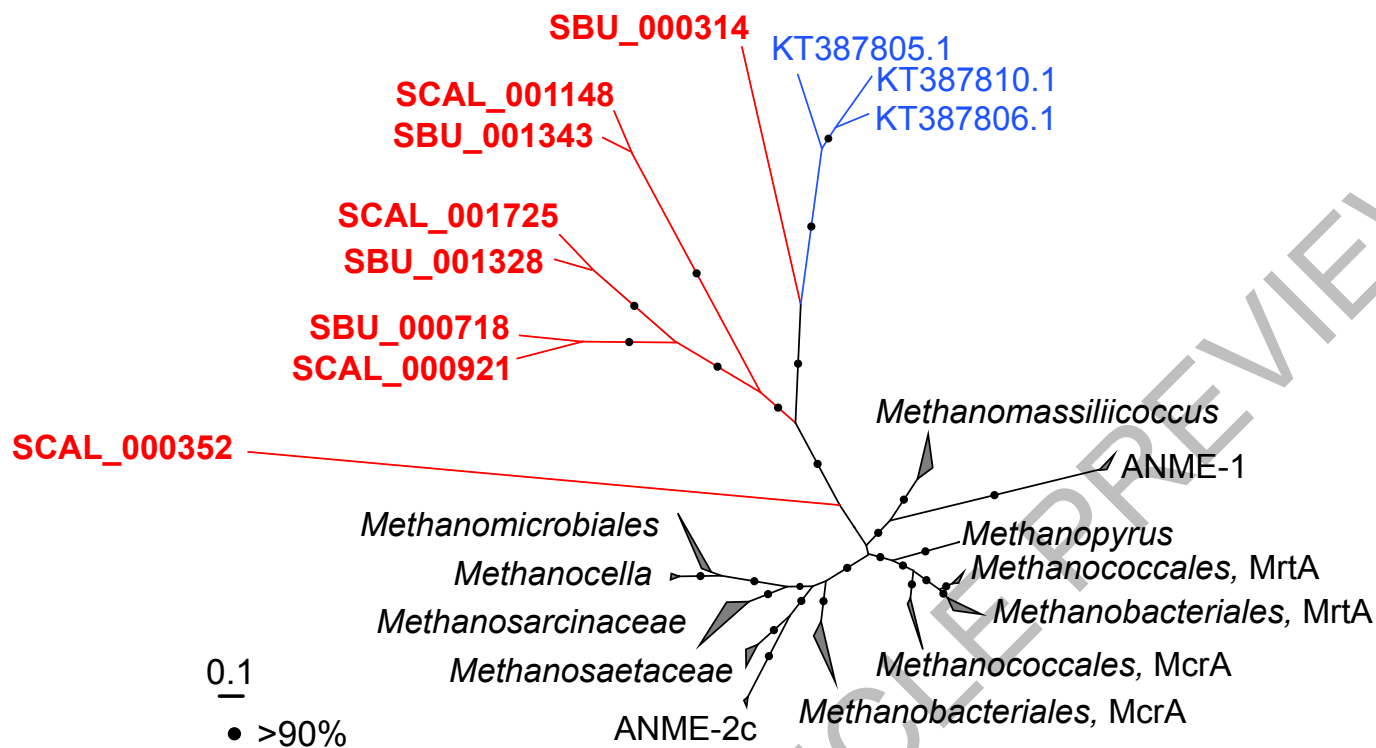
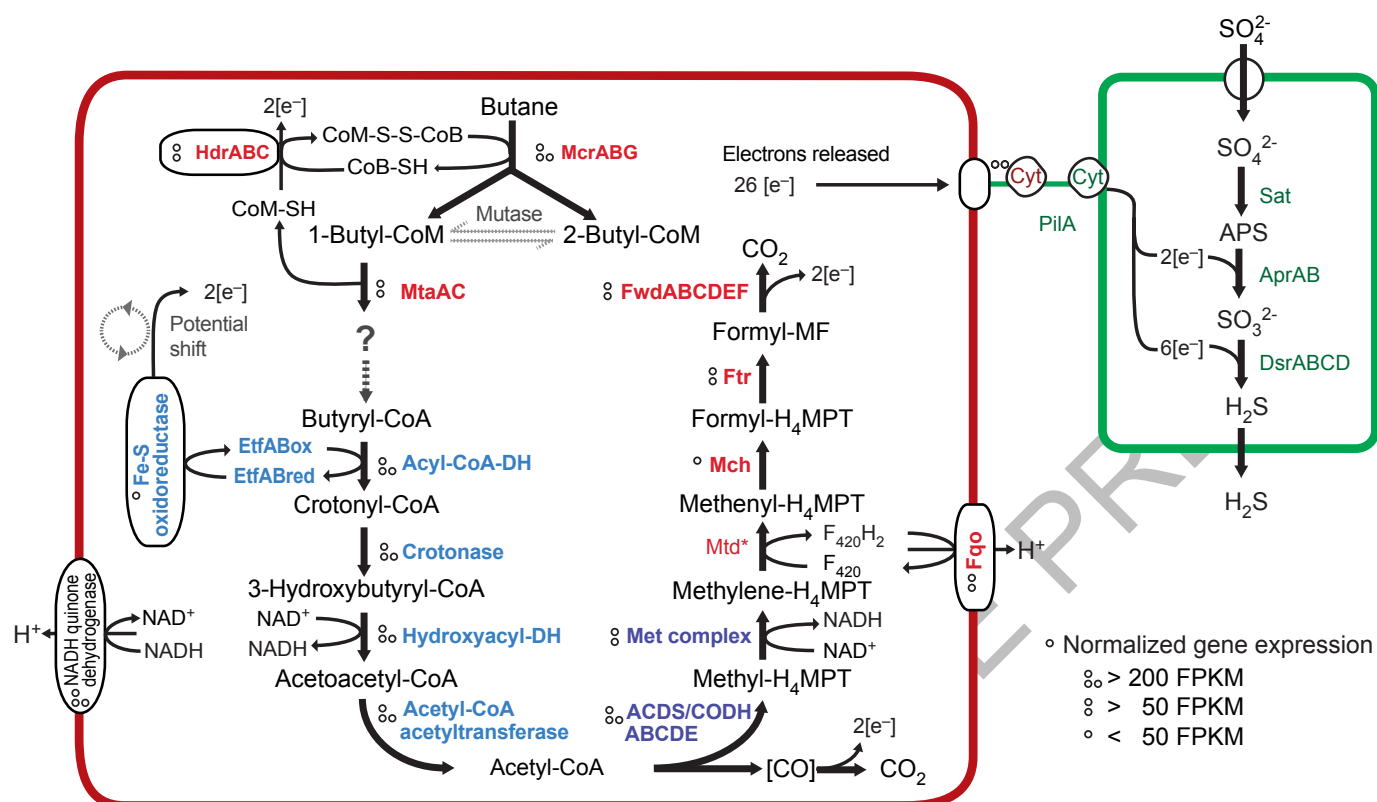


Figure 3 | Phylogenetic affiliation of McrA amino acid sequences present in *Ca. S. butanivorans* and *Ca. S. caldarius*. The phylogenetic tree was constructed based on a maximum likelihood algorithm considering more than 500 amino acid positions. Red branches = sequences from *Ca. Syntrophoarchaeum*; SBU = *Ca. S. butanivorans*; SCAL = *Ca. S. caldarius*, the identifiers refer to the

locus tag of the sequences in the draft genomes. Blue branches indicate *Bathyarchaeota* related sequences from Evans et al²⁹. The scale bar indicates the number of amino acid substitutions per site. Bootstrap values higher than 90% are indicated by filled circles on the corresponding branch.



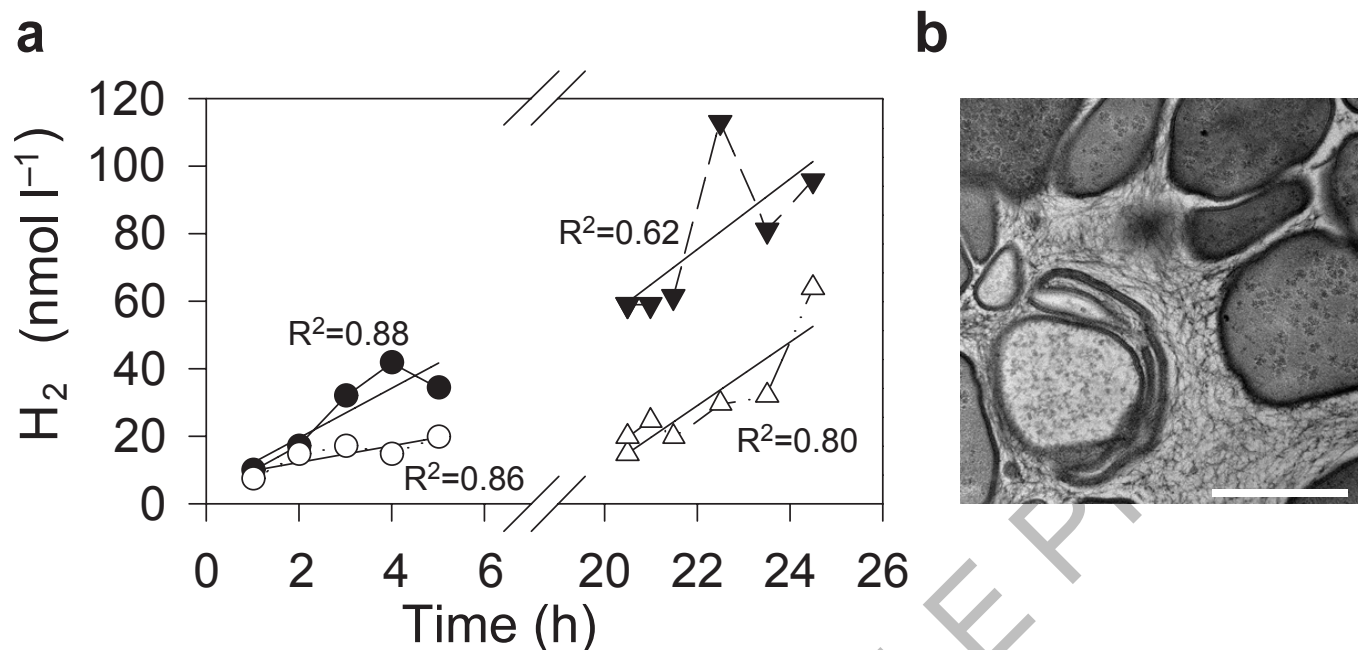


Figure 5 | Testing metabolic interaction of *Ca. Syntrophoarchaeum* and *Ca. D. auxilii* in Butane50 cultures. **a**, Development of hydrogen concentration in replicate active Butane50 cultures ($n=2$) without (circles) and next day continuation with molybdate addition (10 mM final concentration; triangles) to inhibit sulfate reduction. H_2 production was between 3–7 nmol l⁻¹ h⁻¹ and therefore 3–4 orders of magnitude

below a potential hydrogen production of 46,000 nmol l⁻¹ h⁻¹ if H_2 was the intermediate (determined from the measured sulfate reduction rate of 276 μ mol l⁻¹ d⁻¹ before the experiments). **b**, Transmission electron micrograph of an EPON 812 embedded consortium thin-section. The intercellular space in the consortium is filled with abundant nanowire-like structures; scale bar = 0.5 μ m, representative for >30 recorded images.

METHODS

Origin of inoculum and cultivation of the Butane50 enrichment culture. Gas-rich hydrothermally-heated sediments covered with dense mats of *Beggiatoa* were obtained in the Guaymas Basin vent area (27°00.437' N, 11°24.548' W; 2000 m water depth). Samples were collected by push coring using the submersible Alvin (dive 4570) on RV Atlantis during November/December 2009. The sediments were stored anaerobically in butyl rubber stopper-sealed glass vials. In the home laboratory sediments were 1:4 diluted with anoxic artificial seawater (ASW) medium⁴⁹, initially provided with methane as a substrate, and incubated at 50°C. These incubations showed immediate methane-dependent sulfate reduction. After 3 months a subsample was incubated with butane (0.2 MPa). Initially, we did not detect butane-dependent sulfide production, however after 2 months of incubation sulfide production set in. When sulfide concentrations exceeded 15 mM the culture was diluted (1:5) in fresh ASW medium (semi-continuous cultivation) and resupplied with butane. This procedure was repeated several times and resulted in a virtually sediment-free culture after 2 years of cultivation. For quantitative growth experiments, cultures were set up in 150 ml serum bottles containing 80 ml ASW medium, and inoculated with a 20 ml aliquot of a grown culture. Parallel cultures with different starting amounts of butane (5 and 7.5 ml butane in the culture headspace) were prepared. As controls, we used sterile cultures receiving butane, and inoculated cultures lacking butane. All cultures were incubated at 50°C without shaking. Measurements of sulfide production and butane were done in triplicates. **Chemical analyses of sulfide and butane.** Sulfide concentrations were determined by transferring 0.1 ml culture into 4 ml acidified copper sulfate (5 mM) solution. The formation of colloidal copper sulfide was determined photometrically at 480 nm⁵⁰. To quantify butane concentration, volumes of 0.1 ml headspace gas were withdrawn using N₂-flushed, gas-tight syringes. The gas samples were injected without a split into a Shimadzu GC-14B gas chromatograph, equipped with a Supel-Q PLOT column (30 m × 0.53 mm, 30 µm film thickness; Supelco, Bellefonte, USA) and a flame ionization detector. The oven temperature was maintained at 140°C, and the injection and detection temperatures were maintained at 150°C and 280°C, respectively. The carrier phase was N₂ at a flow rate of 3 ml min⁻¹. Samples were analysed in triplicates. Butane concentrations were calculated based on an external calibration curve.

Detection of autofluorescence in Butane50 culture. Consortia from the Butane50 culture were visualized by Confocal Laser Scanning Microscopy (LSM 780; Zeiss, Germany) with an excitation light of 405 nm and an emission filter >463 nm, and by recording the maximum autofluorescence at 470 nm wavelength.

Amplification, sequencing and phylogenetic classification of 16S rRNA and functional genes. Total DNA was extracted from 10 ml of the Butane50 culture pelleted via centrifugation (4000 rpm for 15 min; Eppendorf Centrifuge 5810R) using the FastDNA Spin Kit for Soil (MP Biomedicals) following the manufacturer's protocols. Bacterial and archaeal 16S rRNA gene fragments were amplified using the primer pairs GM3/GM4⁵¹ and Arch20F⁵²/1492R⁵³. Furthermore genes encoding canonical anaerobic hydrocarbon-activating enzymes including *assA/masD* (primer pairs 7757F-1, 7757F-2/8543R⁵⁴) and *bssA* (primer pair 1213F/1987R⁵⁵) were targeted for amplification. For amplification of *assA*, a mixture of forward primers was applied to improve diversity coverage³⁴. Polymerase chain reactions (PCR) were performed in 20 µl volumes containing 0.5 µM of each primer solution, 7.5/6 µg bovine serum albumin solution, 250 µM deoxynucleoside triphosphate (dNTP) mixture, 1 × PCR reaction buffer (5Prime, Germany), 0.25U Taq DNA polymerase (5Prime) and 1 µl DNA template (25–50 ng). PCR reactions (Mastercycler; Eppendorf) included an initial denaturation step of 95°C for 5 min followed by 34 cycles of denaturation (95°C for 1 min.), annealing (1.5 min at 44°C for bacterial 16S primers, or at 58°C for archaeal 16S primers), and extension (72°C for 3 min) and a final 72°C step for 10 min. For amplification of genes encoding canonical hydrocarbon-activating enzymes, the protocol consisted of an initial denaturation step (95°C for 5 min) followed by 34 cycles of denaturation (96°C for 1 min), annealing (58°C for *assA* primers and 55°C for *bssA* primers, both for 1 min) and extension (72°C for 2 min) ending with a final extension (72°C for 10 min). All products were checked on 1% agarose gels, stained with ethidium bromide and visualized with UV light. Amplicons (archaeal and bacterial 16S rRNA gene) were purified (QIAquick PCR Purification Kit; Qiagen) and cloned in *Escherichia coli* (TOPO TA cloning Kit for sequencing; Invitrogen). Clones were screened by standard PCR procedure and positive inserts were sequenced using Taq cycle sequencing with ABI BigDye Terminator chemistry and an ABI377 sequencer (Applied Biosystems, Foster City, CA, USA). Representative full-length sequences were used for phylogenetic analysis using the ARB software package⁵⁶ and the SSURF_NR99_115 SILVA database⁵⁷. Phylogenetic trees of 16S rRNA genes were constructed with RAXML (version 7.7.2) using a 50% similarity filter and the GTRGAMMA model. An extended phylogenetic tree is provided as Supplementary Figure 3. Branch support values were determined using 100 bootstrap

replicates. From the Butane50 culture no *masD/assA* and *bssA* genes could be amplified.

Catalysed reported deposition fluorescence in situ hybridization (CARD-FISH). Cell aliquots were fixed for 2 hours in 2% formaldehyde, washed and stored in phosphate buffered saline (PBS; pH = 7.4); ethanol 1:1. Samples were sonicated (30 s; Sonoplus HD70; Bandelin) and incubated in 0.1 M HCl (1 min) to remove potential carbonate precipitates. Aliquots were filtered on GTTP polycarbonate filters (0.2 µm pore size; Millipore, Darmstadt, Germany). CARD-FISH was performed according to Pernthaler et al.⁵⁸ including the following modifications: cells were permeabilised with a lysozyme solution (0.5 M EDTA pH 8.0, 1 M Tris-HCl pH 8.0, 10 mg ml⁻¹ lysozyme; Sigma-Aldrich) at 37°C for 30 minutes and with a proteinase K solution (0.5 M EDTA, 1 M Tris/HCl, 5 M NaCl, 7.5 µM of proteinase K; Merck, Darmstadt, Germany) for 5 min at room temperature; endogenous peroxidases were inactivated by incubation in a solution of 0.15% H₂O₂ in methanol for 30 min at room temperature. Specific 16S rRNA-targeting oligonucleotide probes used were SYNA-407 and HotSeep-1-1456²⁶, both applied at 20% formamide concentration. SYNA-407 was developed during this project using the probe design tool within the ARB software package to specifically detect *Ca. Syntrophoarchaeum*. The probe is highly specific for *Ca. Syntrophoarchaeum* and has at least one mismatch to non-target group sequences in the current database. The stringency of probe SYNA-407 was experimentally tested on the Butane50 culture using 10% to 40% formamide in the hybridization buffer. The sequence of the probe is: 5' AGTCGACACAGGTGCCGA 3'. Three helpers were necessary: hSYNA-388 (5' ACTCGGAGTCCCCTTATC 3'), hSYNA-369 (5' CACTTGCCTGCATTGTAA 3') and hSYNA-426 (5' TATCCGGACAGTCGACAC 3'). Probes were purchased from Biomers (Ulm, Germany). In case of double hybridization, the peroxidases from the first hybridization were inactivated by incubating the filters in 0.30% H₂O₂ in methanol for 30 min at room temperature. As fluorochromes Alexa Fluor 488 and Alexa Fluor 594 were used. Finally the filters were stained with DAPI (4',6'-diamino-2-phenylindole) and analysed by epifluorescence microscopy (Axiophot II Imaging, Zeiss, Germany). Selected filters were analysed by Confocal Laser Scanning Microscopy (LSM 780, Zeiss, Germany).

Extraction of genomic DNA, library construction and sequencing. Genomic DNA was extracted from 15 ml of the Butane50 culture using the FastDNA Spin Kit for Soil (MP Biomedicals, Illkirch, France). For paired-end library preparation the TruSeq DNA PCR-Free Sample Prep Kit (Illumina) was used including the following modifications of the manufacturer's guidelines: A total amount of 700 ng DNA (in 50 µl volume) was fragmented in 500 µl nebulization buffer (50% glycerol v/v, 35 mM Tris-HCl, 5 mM EDTA), using a Nebulizer (Roche), with a fragmentation time of 3 min, and applied pressure of 32 psi. The fragmented DNA was purified via a MinElute purification column (Qiagen). Following end repair, the first size selection step (removal of large DNA fragments) was done with a sample purification bead/H₂O mixture of 6/5 (v/v).

For mate-pair library construction, genomic DNA was extracted from 35 ml Butane50 culture following the protocol after Zhou et al.⁵⁹ with the following modifications: cells were collected by centrifugation of the culture aliquot (3000 × g for 5 min). The pellet was resuspended in 450 µl of extraction buffer, homogenized in a tissue grinder and the mixture was freeze-thawed three times. Subsequently 1350 µl of fresh extraction buffer and 60 µl of Proteinase K were added. In total, 1370 ng of DNA were obtained and used for mate-pair library construction with the Illumina Nextera Mate Pair Sample Preparation Kit following the manufacturer's guidelines with the following modifications: a total amount of 1.3 µg DNA was used and the fragmentation time was reduced to 15 min. Fragments of lengths between 4 kb and 9 kb were obtained on an agarose gel which were then used for further library preparation. Sequencing of both libraries was performed on a MiSeq 2500 instrument (Illumina; 2 × 300 cycles) using v3 sequencing chemistry. In total 4460548 and 21182518 reads were obtained for the paired-end and mate-pair library respectively.

Read processing, bin assembly and data analysis. The paired-end Illumina reads were quality-trimmed after adaptor and contaminant removal using the bbdut tool in BBMap (V34: <http://sourceforge.net/projects/bbmap>; minimum quality value of 20; minimum read length ≥ 50 bp). Overlapping paired-end reads were merged using bbmerge when overlap exceeded 20 bases without mismatches for reads ≥ 150 bp. The 16S rRNA based phylogenetic composition of the paired-end library was estimated using the software phyloFlash (<https://github.com/HRGV/phyloFlash>), which classifies reads taxonomically by mapping reads against the SSU SILVA 119 database using bbmap. For quantification, only unambiguously mapped reads were counted. For the mate-pair library, junctions, contaminants and external adaptors were removed using bbdut. Afterwards, the reads were quality trimmed (quality value ≥ 20 and minimum sequence length 50 bp). Bulk assembly of processed libraries was done with SPAdes (version 3.5.0⁶⁰) including the

BayesHammer error correction step and using default k-mer size recommended for the read length (21, 33, 55, 77, 99, 127). The resulting scaffolds were analysed and binned using the Metawatt software (version 2.1⁶¹), which analyses the GC content, coverage, open reading frames (ORF) and tetranucleotide pattern for each scaffold. The subsequent binning of the scaffolds was based on three different criteria: highly similar tetranucleotide frequency (98% confidence level), coherent taxonomic classification according to BlastP search of the translated ORFs and similar GC content and read coverage in the metagenome. Using the software RNaMmer⁶², the 16S rRNAs present in the bulk assembly were extracted to classify the different bins of the bulk assembly phylogenetically. Bins corresponding to the GoM-Arch87 group were selected and refined. The refinement started with a mapping of the raw reads (from complete libraries) to the selected bins (with a minimum identity of 90% the first time and 97% the next ones) using the bmap tool from the BMap package. The mapped reads were reassembled using SPAdes (same settings as for the bulk assembly), followed by binning in Metawatt. Contigs smaller than 1 kb were removed from the bin. The mate-pair read mapping information of the bin was used to create connectivity graphs using Cytoscape^{63,64} and to remove poorly-connected contigs. After bin refinement, its completeness was checked using AMPHORA2⁶⁵, which screens for 104 archaeal single copy genes; CheckM⁶⁶, which analyses completeness and contamination based on lineage-specific marker sets, in our case *Euryarchaeota* and tRNAscan⁶⁷, which screens for the different tRNA sequences. The final bins were used as draft genome of *Ca. S. butanivorans* and *Ca. S. caldarius* for automated gene annotation in RAST⁶⁸ and genDB⁶⁹ after gene prediction using Glimmer3.02⁷⁰. After selecting the best annotation for each ORF using the automated annotation tool MicHanThi⁷¹, the GenDB results were visualized using the JCoast frontend⁷². All presented genes were manually curated afterwards.

Analysis of species identity and bacterial electron transfer mechanisms.

A HotSeep-1 bin was retrieved and annotated as described above for *Ca. Syntrophoarchaeum*. To compare our HotSeep-1 bin and the published draft genome of *Ca. D. auxilii* (CP013015), JSpecies1.2.1⁷³ was used, which analyses the average nucleotide identity and the tetranucleotide frequency between two genomes. This method was also used to compare the two genome bins of *Ca. Syntrophoarchaeum*. Furthermore, the two HotSeep-1 strains were compared by checking the identity of the following genes: 16S rRNA, 23S rRNA, sulfate adenyltransferase (*sat*), adenylsulfate reductase subunit alpha (*apr* alpha), adenylsulfate reductase subunit beta (*apr* beta) and dissimilatory sulfite reductase subunit alpha (*dsr* alpha) and of the internal transcribed spacer (ITS) region. To study genes encoding pili and cytochromes of HotSeep-1, genes of interest were identified. This selection was manually curated using Blastp and Pfam search. The subcellular localization of cytochromes was predicted using PSORTb (v3.0.2⁷⁴).

Search for canonical alkyl succinate synthase genes. To search for canonical genes of hydrocarbon oxidation in the metagenome and the bins of *Ca. S. butanivorans* and *Ca. S. caldarius*, a protein database of anaerobic hydrocarbon oxidation genes was constructed. Full-length sequences from hydrocarbon degrading enzymes present in the Uniprot database were combined with recently published *masD* sequences⁷⁵. These enzymes were AssA, BssA, MasD, the alpha subunit from naphthylmethylsuccinate synthase (Nms), the alpha subunit from a ring cleaving hydrolase (BamA), and pyruvate formate lyase (Pfl). The bulk assembly and the *Ca. Syntrophoarchaeum* draft genomes were searched against this database using Blastx with an E-value of 10⁻⁵.

Treatment of Butane50 culture with bromoethanesulfonate. Triplicate Butane50 cultures and duplicates of *Ca. D. auxilii* cultures were grown on their respective substrates (butane or hydrogen). Two active Butane50 cultures were incubated with bromoethanesulfonate (BES, 5 mM final concentration) and as growth control, one culture remained untreated. To check the effect of BES on the bacterial partner alone, hydrogenotrophic grown *Ca. D. auxilii* cultures were also treated with 5 mM of BES. Sulfate-reducing activity was determined by sulfide measurements as described above.

Phylogenetic analysis of methyl-CoM reductases in *Ca. Syntrophoarchaeum* draft genomes. The McrA amino acid sequences in the genomes of *Ca. S. butanivorans* and *Ca. S. caldarius* were extracted from the genomic data, and used for a phylogenetic reconstruction. 124 reference McrA protein sequences longer than 450 amino acids from public databases were aligned with Muscle3.7⁷⁶, accession numbers of these sequences are provided in the Supplementary Table 4. After manual refinement of the alignment a masking filter accounting the alignment ambiguity of each column was designed using the ZORRO software⁶⁵. Phylogenetic trees were calculated using Maximum likelihood algorithm RAXML (version 8.2.6⁷⁷) with the masking filter and the PROTGAMMA model with LG as amino acid substitution model and empirical base frequencies. These were the best-fitting conditions according to RAXML using both Akaike and Bayesian information criterion. To find the optimal tree topology 149 bootstraps were calculated

according to the bootstrap convergence criterion of RAXML. To verify results of the presented phylogenetic affiliation, the phylogenetic analyses were repeated using IQ-TREE⁷⁸ with LG+I+F+C20 as substitution model on the same alignment (Supplementary Figure 1a).

To avoid the possibility of long branch attraction, further partial McrA sequences of *Bathyarchaeota* (Supplementary Table 4) were included and only the McrA sequence regions common between the partial McrAs of *Bathyarchaeota* and our previous set of full-length sequences (>300 residues) was considered for phylogenetic analysis. First, it was confirmed that using these regions for phylogenetic analysis resulted in similar tree topology as using the full-length sequences by calculating a phylogenetic tree using RAXML (PROTGAMMALG+I+F) with the respective parts of all full-length sequences (the data set used in the previous phylogenetic analysis; Supplementary Figure 1b). Then the partial sequences of *Bathyarchaeota* were included into the set to perform phylogenetic analysis of the common McrA sequence parts using both RAXML (PROTGAMMALG+I+F, Supplementary Figure 1c) and IQ-tree (LG+I+F+C20, Supplementary Figure 1d). Finally, to check if the overall tree topology was influenced by the deeply-branching SCAL_000352 sequence, a tree using RAXML (PROTGAMMALG+I+F) with only full-length sequences but excluding the SCAL_000352 sequence was constructed (Supplementary Figure 1e). All resulting trees were plotted using the iTol webserver⁷⁹.

Verification of different McrA operons. To test the correct genome assembly and to confirm the presence of four *mcrA* genes per bin, a *mcrA* clone library was constructed. For each of the 8 *mcrA* genes found in the two *Ca. Syntrophoarchaeum* bins primer sets were developed, which were used for PCR amplification from Butane50 culture DNA (Supplementary Table 5). PCR reactions (20 µl volume) were performed containing 1 µM primer each, 200 µM dNTPs, 1 × PCR buffer, and 0.5 U DNA polymerase (TaKaRa Taq, TaKaRa Bio Europe, France) under the following conditions: initial denaturation at 95 °C for 5 min, followed by 39 cycles of denaturation (96 °C, 1 min), annealing for 1 min, elongation (72 °C, 2 min), and a final elongation step (72 °C, 10 min). For two primer sets, amplification was done with Phusion High-Fidelity DNA Polymerase (Thermo Fischer Scientific, Germany) using 50 µl reactions containing 1.5 mM MgCl₂, 3% (v/v) DMSO, 0.4 µM primer each, 50 µM dNTPs, 1 × PCR buffer, and 1 U DNA polymerase under the following conditions: initial denaturation at 98 °C for 30 s, followed by 39 cycles of denaturation (98 °C, 10 s), annealing for 30 s, elongation (72 °C, 50 s), and a final elongation step (72 °C, 10 min). For annealing temperatures for the individual primer sets see Supplementary Table 5. PCR resulted in multiple bands, therefore amplicons of expected size were excised from an 1% agarose gel and purified using the MinElute Gel extraction kit (Qiagen, Germany). DNA was ligated in a pGEM T-Easy vector (Promega, Madison, WI) and transformed into *E. coli* TOP10 cells (Invitrogen, Carlsbad, CA) according to the manufacturers' recommendations. Sequencing was performed by *Taq* cycle sequencing using a vector-specific primer (M13F or M13R) with a model ABI377 sequencer (Applied Biosystems). Sequence data were analysed with the ARB software package⁸⁰.

Extraction of RNA, library construction and data analysis. Total RNA was extracted from 100 ml of an active Butane50 culture, which was kept at 50 °C during the whole procedure: first most medium (>90%) was replaced by butane gas, whereas the biomass remained at the bottom of the bottle. Then RNA was preserved by adding 90 ml preheated RNAlater (Sigma-Aldrich; 10:1 RNAlater vs. sample) for 1 hour. Subsequently this mixture was filtered through an RNA-free cellulose nitrate filter (pore size 0.45 µm; Sartorius; Göttingen, Germany). The filter was extracted in an RNase-free tube with glass beads and 600 µl of RNA Lysis Buffer (Quick-RNA MiniPrep, Zymoresearch, USA) applying bead beating (2 cycles of 6 m/s for 20 s). The lysate was cleared by centrifugation (10000g; 1 min) and the supernatant was used for RNA extraction with the Quick-RNA MiniPrep Kit (Zymoresearch, Irvine, CA, USA) according to the manufacturer's guidelines but omitting the on-column DNase treatment step. The RNA extract was cleaned from DNA by incubating it at 37 °C for 40 min with 10 µl of DNase I (DNase I recombinant, RNA-free; Roche Diagnostics, Mannheim, Germany), 7 µl of 10 × incubation buffer (Roche) and 2 µl of RNase-Inhibitor (Protector RNase Inhibitor, Roche Diagnostics, Mannheim, Germany). DNases were inactivated by heating for 10 min to 56 °C. Subsequently the RNA was purified with the RNeasy MinElute Cleanup Kit (QIAGEN, Hilden, Germany). In total 450 ng high quality RNA were obtained.

The TruSeq Stranded Total RNA Kit (Illumina) was used for RNA library preparation. The rRNA depletion step was omitted. 80 ng of the total RNA (in 5 µl volume) was combined with 13 µl of 'Fragment, Prime and Finish mix', for the RNA fragmentation step according to the Illumina TruSeq stranded mRNA sample preparation guide. Subsequent steps were performed as described in the sample preparation guide. The library was sequenced on a MiSeq instrument; with v3 sequencing chemistry in 2 × 75 cycles paired-end runs. The resulting reads

were pre-processed including removal of adaptors and contaminants and quality trimming to Q10 using bbdutk v34 from the BBMAP package. Trimmed reads were used to quantify the 16S rRNA gene based phylogenetic composition of the library by phyloFlash as described above for the DNA paired-end library. Trimmed reads were also mapped to the bins of interest (*Ca. S. butanivorans*, HotSeep-1) using bbmap with a minimum identity of 97%. The expression level of each gene was quantified by counting the number of unambiguously mapped reads per gene using featureCount⁸⁰ with the $-p$ option to count fragments instead of reads. To compare expression levels between genes, absolute fragment counts per genes were converted into Fragments Per Kilobase of transcript per Million mapped reads (FPKM⁸¹) as follows:

$$\text{FPKM}_i = \frac{C_i}{L_i \sum_j C_j} \times 10^9$$

where i = any specific gene, j = sum of all the transcribed genes, C = counts and L = length (bp).

Protein analysis by nanoLC-MS/MS. For total protein analysis, the cells from 50 ml of grown (ca. 10 mM sulfide) Butane50 enrichment culture were harvested by centrifugation, frozen in liquid nitrogen and stored at -20°C until analysis. The cell pellets were suspended in 30 μl of 50 mM ammonium bicarbonate buffer, and lysed by three 60 s freeze-thaw cycles between liquid nitrogen and $+40^\circ\text{C}$ (thermal shaker, 1400 rpm). The cell lysate was incubated with 50 mM dithiothreitol at 30°C for 1 h, followed by alkylation with 200 mM iodoacetamide for 1 h at room temperature, in the dark, and trypsin digestion (0.6 μg trypsin, Promega) overnight at 37°C . Peptides were desalted using C18 Zip Tip columns (Millipore), and analysed by nLC-MS/MS using an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a nanoUPLC system (nanoAquity, Waters) as described before⁸².

Peptide identification was conducted by Proteome Discoverer (v1.4.1.14, Thermo Fisher Scientific) using the Mascot search engine with the annotated metagenome of *Ca. Syntrophoarchaeum* as database⁸². Peptides were considered to be identified by Mascot when a probability of 0.05 (probability-based ion score threshold of 40) was achieved. emPAI values calculated by Mascot for identified proteins were used as semi-quantitative measure to estimate the abundance of proteins in the analysed sample⁸³. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium⁸⁴ via the PRIDE partner repository⁸⁵.

Synthesis of authentic standards. To synthesize 1-butyl-CoM and 2-butyl-CoM, 5 g of coenzyme M (Na 2-mercaptoethanesulfonate, purity 98%; Sigma Aldrich) were dissolved in 40 ml of a 30% (v/v) ammonium hydroxide solution, in serum vials. Twice the molar amount of 1-bromobutane (purity 99%; Sigma Aldrich) or 2-bromobutane (purity 98%; Sigma Aldrich) were added, the serum bottles were closed with butyl rubber septa and incubated at room temperature with vigorous shaking (500 rpm) for 4 h. The aqueous phase was separated from the excess hydrophobic 1- or 2-bromobutane via separatory funnels. Residual, dissolved 1- or 2-bromobutane was removed by bubbling with nitrogen. The solutions were analysed for the presence of 1-butyl-CoM or 2-butyl-CoM by FT-ICR-MS analysis without further purification. Both solutions contained a major m/z peak at 197.0311; no m/z peaks indicative of free CoM, CoM dimers, 1- or 2-bromobutane were detected. Both standards were stable and no interconversion of isomers was observed.

Metabolite extraction. For preparation of cell extracts, volumes of 20 ml were collected from grown Butane50 cultures (sulfide concentrations of 14–15 mM) under anoxic conditions. The cells were harvested by centrifugation (10 min, 10,000 rpm, 4°C), washed twice with a 100 mM ammonium bicarbonate solution, and finally suspended in 1 ml of acetonitrile: methanol: water solution (40:40:20 v/v). Glass beads (0.1 mm diameter, Roth) were added (0.3 g per tube), and the cells were lysed with a PowerLyzer 24 bench top bead-based homogenizer (MO BIO Laboratories, Carlsbad, CA) using 5 cycles of 2000 rpm for 50 sec, with a 15 sec pause between cycles. Prior use, the glass beads were treated with 1N HCl solution and washed twice with deionized water. Glass beads and cell debris were removed by centrifugation, and the aqueous cell extracts were stored in glass vials at 4°C until analysis.

Mass spectrometry of cell extracts and standards. Authentic standards and cell extract samples were measured with ultra-high resolution mass spectrometry (Solarix XR 12T Fourier transform ion cyclotron resonance mass spectrometer, Bruker Daltonics Inc., Billerica, MA) with negative electrospray ionization (capillary voltage: 4.5 kV) in direct infusion mode (4 $\mu\text{l}/\text{min}$ and 0.1 s accumulation time). Spectra were recorded with a 2 MWord time domain (0.42 s transient length) between m/z 74 and 3000 resulting in a mass resolution of ca. 250,000 at m/z 200. Instrument mass accuracy was linearly calibrated with low-molecular mass fatty acids (C4–C12) between 88 and 199 Da, resulting in an average root-mean square error of the calibration masses of 39 ppb ($n = 7$). For each measurement, 64 (Butane50 samples), or 128 (controls) spectra were co-added (lock mass: 143.10775 m/z) and internally recalibrated with naturally present fatty acids.

Collision induced fragmentation of m/z 197 was carried out after quadrupole isolation (10 Da window) with 12 V collision energy and 128 scans per measurement (lock mass: 199.17035 m/z). The 1-butyl-CoM and 2-butyl-CoM standards were diluted to ca. 10 $\mu\text{g}/\text{ml}$ and checked for appropriate collision energy and fragment pattern. Fragment masses 89.0430 ($\text{C}_4\text{H}_9\text{S}^-$) and 80.9652 (HSO_3^-) were then used as indicative fragment for butyl-CoM in the cell extracts. The formation of an even-electron fragment HSO_3^- from bisulfite is favoured when a β -H-atom is present⁸⁶. However, SO_3^- ($m/z = 79.9674$) was also produced upon fragmentation of the standards.

Fragmentation information of the butyl-CoM standards was used to implement a UPLC-MS/MS method to validate the isomeric form of m/z 197.031 in the samples. A triple quadrupole mass spectrometer (Xevo TQ-S, Waters Cooperation, Manchester, UK) in negative electrospray ionization mode was used in multiple reaction monitoring (MRM) mode. Indicative butyl-CoM transitions (m/z 197 > 89 and m/z 197 > 81) were initially optimized (cone voltage and collision energy) by direct infusion of standard solutions into the mass spectrometer. The mass spectrometer was coupled to a UPLC (ACQUITY I-Class, Waters Cooperation Milford, MA, USA) equipped with a reversed phase column (HSS T3, 25 cm, Waters) and run with a binary gradient (1% methanol in Water to 90% methanol) at a flow rate of 0.3 ml min^{-1} . For each analysis, 10 μl were injected into the UPLC. Retention time, presence of both MRM transitions and relative ion ratios as compared to the standards were used as quality criteria.

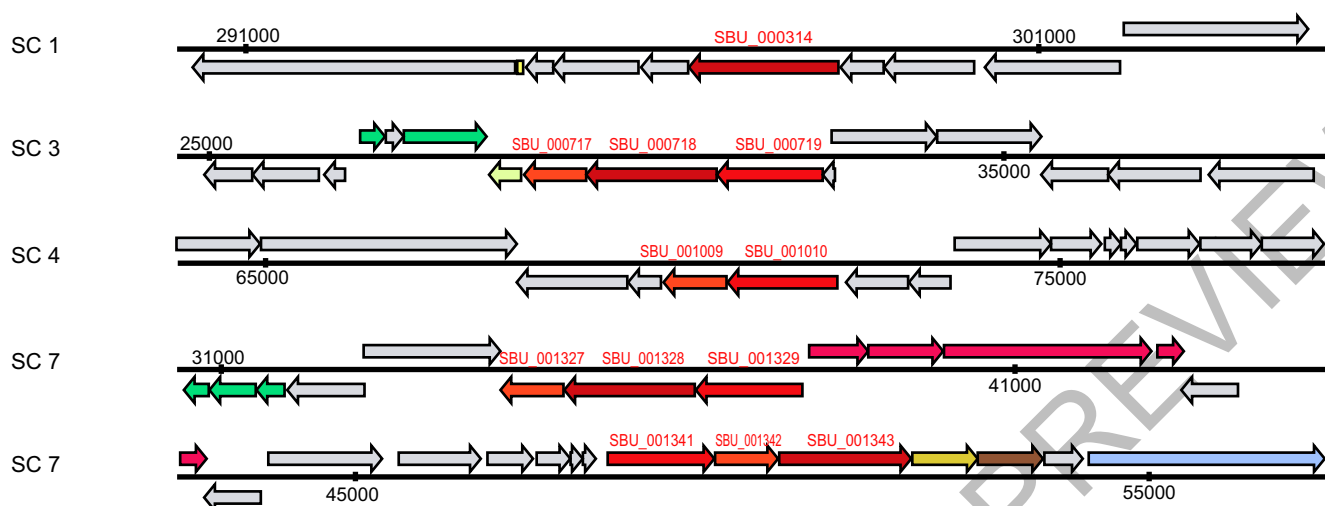
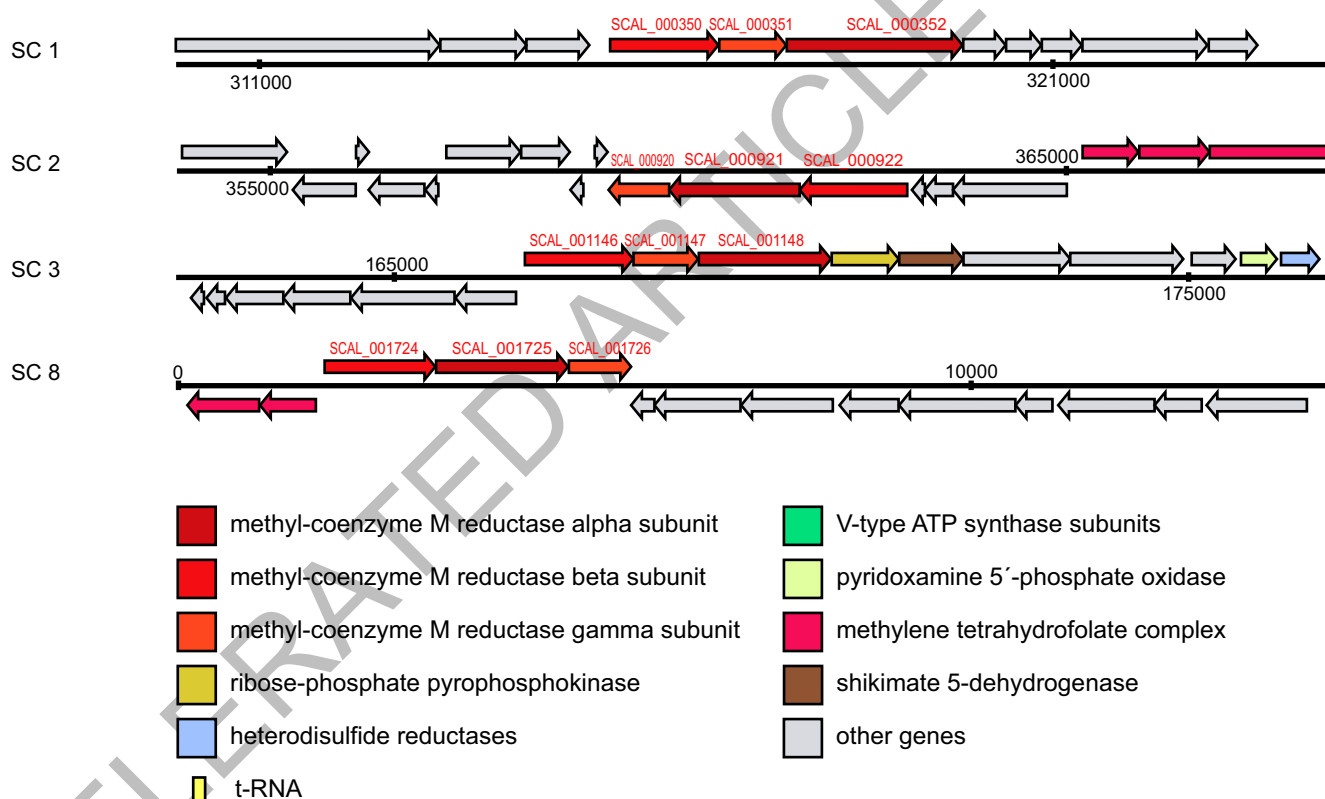
Quantification of hydrogen production in experiment. Hydrogen production in the Butane50 culture was measured by analysing the headspace of replicate incubations which were constantly agitated on a shaking table in a 50°C -incubator. The butane-dependent sulfide production (and therefore potential hydrogen production) was determined from tracking the sulfide production (as above) for 4 weeks. Gas phase (1 ml) was sampled with a gas-tight syringe to determine hydrogen concentrations (i) before changing the headspace, (ii) after exchanging the headspace in 30 minute intervals for 6 hours (iii) the next day, before and after addition of sodium molybdate solution (10 mM final concentration) to the culture to stop potential hydrogen-dependent sulfate reduction. Gas phase was immediately injected into a Peak Performer 1 gas chromatograph (Peak Laboratories, Palo Alto, CA) equipped with a reducing compound photometer. Development of hydrogen concentrations were converted into hydrogen production rates and compared with potential hydrogen production rates according to a stoichiometry of 4:1 (H_2 -production vs. sulfate reduction).

Transmission electron microscopy. A 100 ml grown Butane50 culture was concentrated by centrifugation at 2,000 rpm using a Stat Spin Microprep 2 table-top centrifuge. Aliquots were placed in aluminium platelets of 150 μm depth containing 1-hexadecan⁸⁷. The platelets were frozen using a Leica EM HPM100 high pressure freezer (Leica Mikrosysteme, Wetzlar, Germany). The frozen samples were transferred to an Automatic Freeze Substitution Unit (Leica EM AFS2) and substituted at -90°C in a solution containing anhydrous acetone, 0.1% tannic acid for 24 h and in anhydrous acetone, 2% OsO_4 , 0.5% anhydrous glutaraldehyde (Electron Microscopy Sciences, Ft. Washington, USA) for additional 8 h. After a further incubation over 20 h at -20°C samples were warmed up to $+4^\circ\text{C}$ and washed with anhydrous acetone subsequently. The samples were embedded at room temperature in Agar 100 (Epon 812 equivalent) at 60°C over 24 h. Thin sections (80 nm) were examined using a Philips CM 120 BioTwin transmission electron microscope (Philips Inc. Eindhoven, The Netherlands). Images were recorded with a TemCam F416 CMOS camera (TVIPS, Gauting, Germany), for additional images see Supplementary Figure 4.

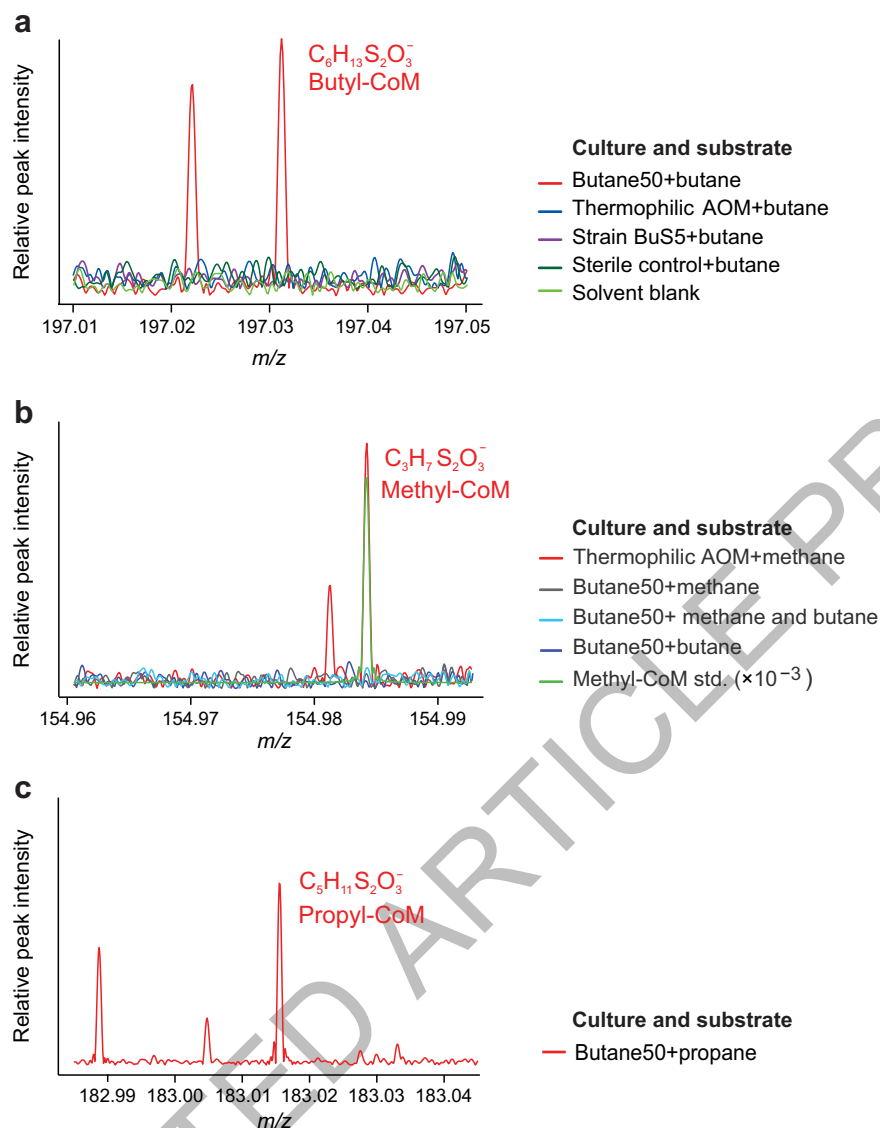
Data availability. All sequence data are archived in NCBI database under the BioSample number SAMN05004607. Representative full-length 16S rRNA gene sequences of the clone library of the Butane50 culture have been submitted to NCBI under accession numbers KX812780–KX812802. Draft genomes of the *Ca. Syntrophoarchaeum* organisms can be found under the BioProject PRJNA318983 (*Ca. S. butanivorans*) and PRJNA319143 (*Ca. S. caldarius*). Metagenomic and metatranscriptomic reads have been submitted to the short read archive SRS1505411. The mass spectra of the proteomic dataset have been deposited to the ProteomeXchange Consortium with the dataset identifier PXD005038.

49. Widdel, F. & Bak, F. In *The prokaryotes* Vol. 4 (ed Trüper HG Balows A, Dworkin M, Harder W, Schleifer KH) 3352–3378 (Springer, 1992).
50. Cord-Ruwisch, R. A quick method for the determination of dissolved and precipitated sulfides in cultures of sulfate-reducing bacteria. *Microbiol Meth* **4**, 33–36 (1985).
51. Muyzer, G., Teske, A., Wirsén, C. O. & Jannasch, H. W. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164**, 165–172 (1995).
52. Massana, R., Murray, A. E., Preston, C. M. & DeLong, E. F. Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Applied and environmental microbiology* **63**, 50–56 (1997).

53. Teske, A. *et al.* Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Applied and environmental microbiology* **68**, 1994–2007 (2002).
54. von Netzer, F. *et al.* Enhanced gene detection assays for fumarate-adding enzymes allow uncovering of anaerobic hydrocarbon degraders in terrestrial and marine systems. *Appl. Environ. Microbiol.* **79**, 543–552 (2013).
55. Callaghan, A. V. *et al.* Diversity of benzyl- and alkylsuccinate synthase genes in hydrocarbon-impacted environments and enrichment cultures. *Environmental science & technology* **44**, 7287–7294 (2010).
56. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Research* **32**, 1363–1371, doi:10.1093/nar/gkh293 (2004).
57. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–596, doi:10.1093/nar/gks1219 (2013).
58. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and environmental microbiology* **68**, 3094–3101 (2002).
59. Zhou, J., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Applied and environmental microbiology* **62**, 316–322 (1996).
60. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).
61. Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3**, 410, doi:10.3389/fmicb.2012.00410 (2012).
62. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research* **35**, 3100–3108 (2007).
63. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
64. Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature biotechnology* **31**, 533–538 (2013).
65. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**, 1033–1034 (2012).
66. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Report No. 2167-9843, (PeerJ PrePrints, 2014).
67. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research* **33**, W686–W689 (2005).
68. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**, 75 (2008).
69. Meyer, F. *et al.* GenDB - an open source genome annotation system for prokaryote genomes. *Nucleic acids research* **31**, 2187–2195, doi:10.1093/nar/gkg312 (2003).
70. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).
71. Quast, C. MicHanThi-design and implementation of a system for the prediction of gene functions in genome annotation projects. *Universität Bremen, Diploma work* (2006).
72. Richter, M. *et al.* JCoast-A biologist-centric software tool for data mining and comparison of prokaryotic (meta) genomes. *BMC bioinformatics* **9**, 177 (2008).
73. Richter, M. & Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* **106**, 19126–19131, doi:10.1073/pnas.0906412106 (2009).
74. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615, doi:10.1093/bioinformatics/btq249 (2010).
75. Stagars, M. H., Ruff, S. E., Amann, R. & Knittel, K. High diversity of anaerobic alkane-degrading microbial communities in marine seep sediments based on (1-methylalkyl)succinate synthase genes. *Frontiers in Microbiology* **6**, doi:10.3389/fmicb.2015.01511 (2016).
76. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).
77. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
78. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
79. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274, doi:10.1093/molbev/msu300 (2015).
80. Liao, Y., Smyth, G. & Shi, W. featureCounts: an efficient general-purpose read summarization program. *arXiv* **1305**, 16 (2013).
81. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Meth.* **5**, 621–628 (2008).
82. Wagner, A., Cooper, M., Ferdi, S., Seifert, J. & Adrian, L. Growth of *Dehalococcoides mccartyi* strain CBDB1 by reductive dehalogenation of brominated benzenes to benzene. *Environmental science & technology* **46**, 8960–8968, doi:10.1021/es3003519 (2012).
83. Ishihama, Y. *et al.* Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Molecular & Cellular Proteomics* **4**, 1265–1272, doi:10.1074/mcp.M500061-MCP200 (2005).
84. Vizcaino, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotech* **32**, 223–226, doi:10.1038/nbt.2839 (2014).
85. Vizcaino, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic acids research* **44**, D447–D456, doi:10.1093/nar/gkv1145 (2016).
86. Jariwala, F. B., Wood, R. E., Nishshanka, U. & Attygalle, A. B. Formation of the bisulfite anion (HSO₃⁻, m/z 81) upon collision induced dissociation of anions derived from organic sulfonic acids. *Journal of Mass Spectrometry* **47**, 529–538 (2012).
87. Studer, D., Michel, M. & Müller, M. High pressure freezing comes of age. *Scanning microscopy* **3**, 253–268 (1988).

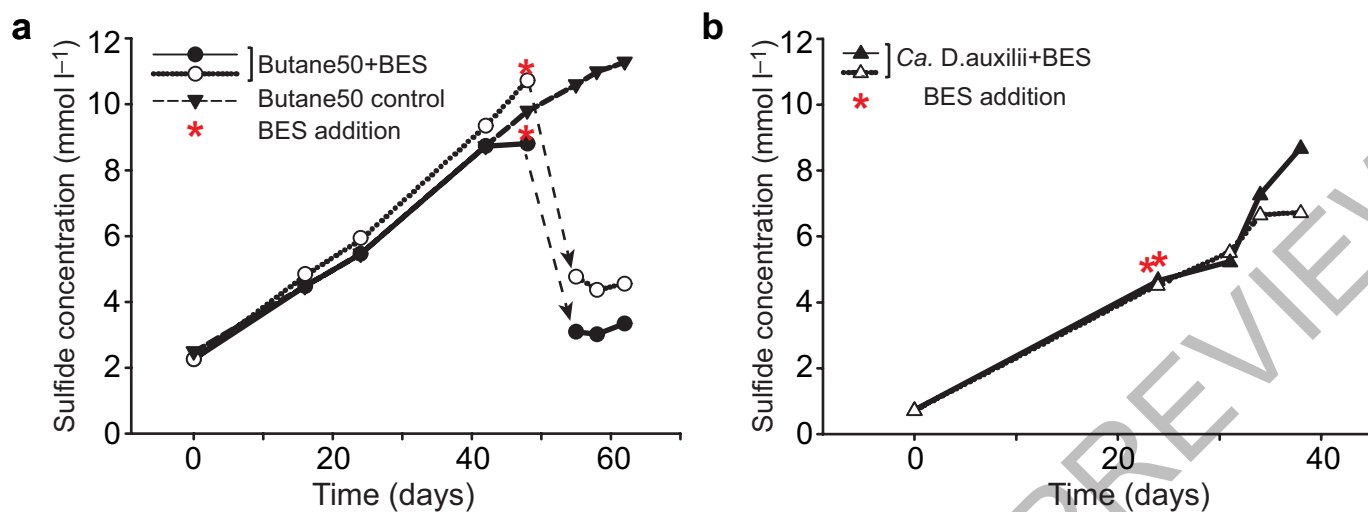
Ca. S. butanivorans* (SBU)**Ca. S. caldarius* (SCAL)**

Extended Data Figure 1 | Genetic structure of *mcr* genes in *Ca. Syntrophoarchaeum*. In *Ca. S. butanivorans* one *mcr* gene set is separated, with the *mcrA* subunit in scaffold 1, and *mcrB* and *mcrG* in scaffold 4.



Extended Data Figure 2 | Experiments validating production of alkyl-CoM compounds in anaerobic cultures. **a**, Screening for butyl-CoM in Butane50, in a thermophilic AOM culture supplied with butane ($n = 2$ with 2 different sampling time points), in BuS5 cultures ($n = 3$) and in controls. The mass peak of butyl-CoM ($m/z = 197.0312$) was only found in the Butane50 culture. **b**, Screening for methyl-CoM ($m/z = 154.984$; mass accuracy -0.15 ppm) in the thermophilic AOM culture supplied with

methane ($n = 3$) and in the Butane50 culture ($n = 2$). Methyl-CoM was only found in the thermophilic AOM culture. These analyses (**a**,**b**) indicate high substrate specificity of the organisms in the two different cultures. **c**, The also established propane-degrading culture ($n = 1$) forms propyl-CoM ($m/z = 183.016$; mass accuracy -0.21 ppm) showing CoM-type activation of this substrate.



Extended Data Figure 3 | Effect of bromoethanesulfonate (BES) on Butane50 and *Ca. D. auxilii* cultures. After growth on their specific substrates, BES (5 mM final concentration) was added to **a**, Butane50 cultures (circles, $n = 2$) where it immediately inhibited butane-dependent

sulfate reduction compared to a control without BES (triangles, $n = 1$) and to **b**, *Ca. D. auxilii* cultures (triangles, $n = 2$) where it had no influence on hydrogen-dependent sulfate reduction.

Extended Data Table 1 | Microbial diversity in the early AOM enrichment²¹ used as inoculum and in the Butane50 culture

Phylogenetic group	Early AOM enrichment		Butane50 culture*	
	Clones	Clones	Metagenome reads	Metatranscriptome reads
Archaea				
<i>Euryarchaeota</i>				
<i>Methanomicrobia</i>				
GoM-Arch87		46	596	32009
<i>Methermicoccus</i>				2316
ANME-1	46			1465
<i>Thermoplasmata</i>	6			
19c-33 cluster		38	199	
pMC2A24 cluster		1	30	
20c-4 cluster			132	
<i>Halobacteria</i>	4			
Sum	56	85	957	35790
Bacteria				
<i>Proteobacteria</i>				
<i>Deltaproteobacteria</i>				
HotSeep-1 cluster	53	63	575	4780
<i>Syntrophobacteraceae</i>	1	5		626
Others	3	3		
<i>Spirochaete</i>				
Kazan-3B-09		1		
TA06		8	44	
Candidate division OP3	3	6	36	
Candidate division KB1			159	827
Candidate division WS3			33	
BHI80-139		1		
<i>Chloroflexi</i>	3			
<i>Anaerolineaceae</i>		3	138	
Others	17			
Sum	80	92	985	6233

*based on 16S rRNA gene sequences retrieved by clone library approach from the Butane50 culture, and found in the metagenome and metatranscriptome libraries using the phyloFlash software. Taxa which account for $\geq 1\%$ of all 16S rRNA gene sequences are shown.

Extended Data Table 2 | Draft genome information and pairwise comparison of whole genome identity of *Ca. S. butanivorans* and the *Ca. S. caldarius*

	<i>Candidatus</i> Syntrophoarchaeum butanivorans	<i>Candidatus</i> Syntrophoarchaeum caldarius	
Size (base pair)	1,456,963	1,666,081	
Scaffolds/Contigs	16/21	10/15	
Scaffold N50 (bp)	219,218	410,601	
Coverage (times)	360	134	
GC content (%)	48.7	45.5	
Number of ORFs	1,604	1,790	
rRNAs	3	3	
tRNAs	39	44	
Genome Completeness (%)	85 ¹ ; 88 ² ; 89 ³	95 ¹ ; 97 ² ; 96 ³	
Contamination ³ (%)	0.97	0.32	
Duplication of single copy genes ³	2	1	
Strain heterogeneity ³ (%)	0	0	
	Average Nucleotide Identity (Blast)	Average Nucleotide Identity (MUMmer)	Tetranucleotide frequency
Ca. S. butanivorans/Ca. S. caldarius	73.57	89.57	88.19
Ca. S. caldarius/Ca. S. butanivorans	73.56	89.03	88.19

¹Based on tRNA completeness, using tRNAscan;²based on archaea-specific single copy genes, using AMPHORA2; ³based on lineage-specific marker genes of *Euryarchaeota*, using CheckM

Extended Data Table 3 | Genes encoding enzymes for butane activation, candidates for further conversion reactions and butyryl-CoA oxidation in *Ca. S. butanivorans*

Gene	Feature	Locus tag	Transcriptome		Proteome
			Absolute reads	FPKM	emPAI
Butane activation					
mcrA	Methyl coenzyme M reductase alpha	SBU_000314	8320	491.4	0.15
mcrA	Methyl coenzyme M reductase alpha	SBU_000718	1640	108.9	0.05
mcrA	Methyl coenzyme M reductase alpha	SBU_001328	514	34.13	
mcrA	Methyl coenzyme M reductase alpha	SBU_001343	1077	70.5	0.11
mcrB	Methyl coenzyme M reductase beta	SBU_000719	5047	418.0	0.07
mcrB	Methyl coenzyme M reductase beta	SBU_001010	2628	210.1	0.46
mcrB	Methyl coenzyme M reductase beta	SBU_001329	311	25.7	
mcrB	Methyl coenzyme M reductase beta	SBU_001341	602	49.0	0.3
mcrG	Methyl coenzyme M reductase gamma	SBU_000717	689	97.7	
mcrG	Methyl coenzyme M reductase gamma	SBU_001009	1787	243.0	0.5
mcrG	Methyl coenzyme M reductase gamma	SBU_001327	642	91.4	
mcrG	Methyl coenzyme M reductase gamma	SBU_001342	366	52.3	0.37
hdrA	CoB--CoM heterodisulphide reductase,	SBU_000296	2586	94.9	0.12
hdrB	CoB--CoM heterodisulphide reductase,	SBU_000294	222	27.4	
hdrC	CoB--CoM heterodisulphide reductase,	SBU_000295	127	36.3	0.35
hdrA	Heterodisulfide reductase subunit A	SBU_001347	459	16.9	
hdrA	Heterodisulfide reductase subunit A	SBU_001502	903	60.7	0.05
hdrA	Heterodisulfide reductase subunit A	SBU_001503	995	56.9	0.14
Candidates for conversion to butyryl-CoA					
mtaA	Methyltransferase corrinoid activation protein	SBU_000376	1503	83.6	0.36
mtaA	Methylcobamide:CoM methyltransferase	SBU_000378	583	61.7	0.09
mtaA	Methylcobamide:CoM methyltransferase	SBU_000450	988	101.7	0.61
mtaA	Methylcobamide:CoM methyltransferase	SBU_001175	274	28.4	
mtaA	Methylcobamide:CoM methyltransferase	SBU_001379	133	13.6	
mtaA	Methylcobamide:CoM methyltransferase	SBU_001480	553	57.7	0.08
mtaC	Corrinoid protein	SBU_000377	1016	156.7	1.33
mtaC	Corrinoid methyltransferase	SBU_001174	95	16.0	
Butyryl-CoA oxidation					
crt	Acyl-CoA dehydrogenase domain-containing	SBU_000172	1751	168.5	1.43
	Acyl-CoA dehydrogenase	SBU_000399	2844	205.1	0.55
	Acyl-CoA dehydrogenase domain-containing	SBU_000724	449	42.9	0.25
	Acyl-CoA dehydrogenase	SBU_001146	177	12.4	0.11
	Crotonase	SBU_000400	1745	236.4	1.65
	3-Hydroxyacyl-CoA dehydrogenase	SBU_000288	2425	312.5	0.81
	3-Hydroxyacyl-CoA dehydrogenase	SBU_000843	1151	162.5	0.12
	Acetyl-CoA acetyltransferase	SBU_000329	103	9.8	
	Acetyl-CoA acetyltransferase	SBU_000402	3198	295.4	0.34
	Acetyl-CoA acetyltransferase	SBU_000404	1644	152.6	0.16
	Acetyl-CoA acetyltransferase	SBU_001291	161	15.5	
etfA	Electron transfer flavoprotein subunit alpha	SBU_000173	557	63.6	0.1
etfB	Electron transfer flavoprotein subunit beta	SBU_000174	506	69.8	0.58
	Fe-S Oxidoreductase	SBU_000175	402	38.8	0.55

Expression as absolute read counts and as fragments per kilobase of transcript per million mapped reads (FPKM) is shown, as well as the corresponding protein abundance as emPAI index.

Extended Data Table 4 | Genes encoding enzymes of C-1 pathway in *Ca. S. butanivorans*

Gene	Feature	Locus_tag	Transcriptome		Proteome
			Absolute reads	FPKM	emPAI
<i>cdhA</i>	Acetyl-CoA decarboxylase/synthase complex alpha	SBU_000891	3756	172.8	0.65
<i>cdhA</i>	Acetyl-CoA decarboxylase/synthase complex alpha	SBU_001568	1518	70.5	0.05
<i>cdhB</i>	Acetyl-CoA decarboxylase/synthase complex epsilon	SBU_000890	557	120.8	0.64
<i>cdhB</i>	Acetyl-CoA decarboxylase/synthase complex epsilon	SBU_001569	325	65.1	0.35
<i>cdhC</i>	Acetyl-CoA decarboxylase/synthase complex beta	SBU_000889	4798	350.2	0.96
<i>cdhC</i>	Acetyl-CoA decarboxylase/synthase complex beta	SBU_001570	2738	212.9	0.09
<i>cdhD</i>	Acetyl-CoA decarboxylase/synthase complex delta	SBU_000887	3183	232.8	3.03
<i>cdhD</i>	Acetyl-CoA decarboxylase/synthase complex delta	SBU_001572	940	79.6	1.41
<i>cdhE</i>	Acetyl-CoA decarboxylase/synthase complex gamma	SBU_000886	3419	266.4	0.52
<i>cdhE</i>	Acetyl-CoA decarboxylase/synthase complex gamma	SBU_001573	1503	118.6	1.21
<i>metV</i>	5,10-Methylenetetrahydrofolate reductase, C-terminal	SBU_000428	220	36.2	
<i>metF</i>	5,10-Methylenetetrahydrofolate reductase	SBU_000429	487	59.8	0.01
<i>hdrA</i>	Heterodisulfide reductase subunit A	SBU_000430	1435	80.5	
<i>hdrB</i>	Heterodisulfide reductase	SBU_000431	703	45.3	
<i>hdrC</i>	Heterodisulfide reductase	SBU_000432	1259	81.5	
<i>mvhD</i>	Methyl-viologen-reducing hydrogenase delta subunit	SBU_000433	356	53.9	
<i>hdrA</i>	Heterodisulfide reductase subunit A	SBU_000434	839	47.2	
<i>mvhD</i>	Methyl-viologen-reducing hydrogenase delta subunit	SBU_000435	585	154	
<i>metV</i>	5,10-Methylenetetrahydrofolate reductase, C-terminal	SBU_001330	263	39.4	0.25
<i>metF</i>	5,10-Methylenetetrahydrofolate reductase	SBU_001331	711	83.9	0.20
<i>hdr</i>	Heterodisulfide reductase	SBU_001332	392	16.5	
<i>mvhD</i>	Methyl-viologen-reducing hydrogenase delta subunit	SBU_001333	170	57.4	
<i>mch</i>	N(5)N(10)-MethenylH4MPT cyclohydrolase	SBU_000838	197	22.6	0.10
<i>ptr</i>	FormylMF-H4MPT formyltransferase	SBU_001141	416	51.4	0.10
<i>fwdA</i>	Formylmethanofuran dehydrogenase subunit A	SBU_000443	1215	79.2	0.11
<i>fwdB</i>	Formylmethanofuran dehydrogenase subunit B	SBU_000444	555	47.3	
<i>fwdB</i>	Formylmethanofuran dehydrogenase subunit B	SBU_000048	226	21.4	
<i>fwdC</i>	Formylmethanofuran dehydrogenase subunit C	SBU_000442	567	81.3	0.12
<i>fwdD</i>	Formylmethanofuran dehydrogenase subunit D	SBU_000047	75	20.6	
<i>fwdD</i>	Formylmethanofuran dehydrogenase subunit D	SBU_000445	151	43.8	
<i>fwdE</i>	Formylmethanofuran dehydrogenase subunit E	SBU_000903	23	3.8	
<i>fwdF</i>	Formylmethanofuran dehydrogenase subunit F	SBU_001540	91	8.7	

Expression as absolute read counts and as fragments per kilobase of transcript per million mapped reads (FPKM) is shown, as well as the corresponding protein abundance as emPAI index.

Extended Data Table 5 | Genes encoding proteins related to electron cycling and energy transfer in *Ca. S. butanivorans*

Gene	Feature	Locus_tag	Transcriptome		Proteome
			Absolute reads	FPKM	emPAI
	[Ni-Fe]-Hydrogenase large subunit	SBU_000461	5409	343.7	0.28
	[Ni-Fe]-Hydrogenase small subunit	SBU_000462	1702	196.6	
	Cytochrome c-type protein	SBU_000189	324	73.2	
	Cytochrome c	SBU_000960	1161	125.5	0.17
	Cytochrome C	SBU_001187	184	15.3	
	Cytochrome c	SBU_001594	62	5.0	0.21
	Multiheme cytochrome	SBU_000341	196	25.4	
	Multiheme cytochrome	SBU_000342	148	13.1	
	Multiheme cytochrome	SBU_000614	646	83.2	
	Multiheme cytochrome	SBU_000694	4224	659.7	
	Multiheme cytochrome	SBU_000777	3660	571.6	
	Multiheme cytochrome	SBU_000778	3189	395.7	
	Multiheme cytochrome	SBU_001337	1953	389.1	0.61
<i>hdrA</i>	Heterodisulphide reductase, subunit A*	SBU_000297	646	27.0	0.40
<i>hdrB</i>	Heterodisulphide reductase, subunit B*	SBU_000298	1146	146.1	
<i>hdrC</i>	Heterodisulphide reductase, subunit C*	SBU_000299	503	91.7	0.31
<i>fdhB</i>	Formate dehydrogenase subunit beta	SBU_000300	838	86.2	0.89
<i>mvhD</i>	Methyl-viologen-reducing hydrogenase delta	SBU_000301	205	53.9	0.78
<i>fqoJ</i>	F420H2:quinone oxidoreductase subunit J	SBU_000209	3	0.6	
<i>fqoK</i>	F420H2:quinone oxidoreductase subunit K	SBU_000210	12	4.1	
<i>fqoL</i>	F420H2:quinone oxidoreductase subunit L	SBU_000211	133	9.7	
<i>fqoM</i>	F420H2:quinone oxidoreductase subunit M	SBU_000213	166	10.6	
<i>fqoN</i>	F420H2:quinone oxidoreductase subunit N	SBU_000214	142	12.7	
<i>fqoA</i>	F420H2:quinone oxidoreductase subunit A	SBU_000215	85	26.7	0.26
<i>fqoBCD</i>	F420H2:quinone oxidoreductase subunit	SBU_000216	323	15.5	0.04
<i>fqoH</i>	F420H2:quinone oxidoreductase subunit H	SBU_000217	171	17.3	
<i>fqoI</i>	F420H2:quinone oxidoreductase subunit I	SBU_000218	112	16.9	
<i>fqoF</i>	F420H2:quinone oxidoreductase subunit F	SBU_000219	94	10.0	
<i>nuoH</i>	NADH:quinone oxidoreductase subunit H	SBU_000563	446	47.5	0.18
<i>nuoD</i>	NADH:quinone oxidoreductase subunit D	SBU_000564	458	45.7	0.26
<i>nuoC</i>	NADH:quinone oxidoreductase subunit C	SBU_000565	236	56.8	0.69
<i>nuoB</i>	NADH:quinone oxidoreductase subunit B	SBU_000566	345	34.2	
<i>nuoA</i>	NADH:quinone oxidoreductase subunit A	SBU_000567	30	9.4	
<i>nuoI</i>	NADH:quinone oxidoreductase subunit I	SBU_000874	119	34.2	0.23
<i>nuoJ</i>	NADH:quinone oxidoreductase subunit J	SBU_000875	67	30.4	
<i>nuoK</i>	NADH:quinone oxidoreductase subunit K	SBU_000877	69	24.4	
<i>nuoL</i>	NADH:quinone oxidoreductase subunit L	SBU_000878	950	54.5	
<i>nuoM</i>	NADH:quinone oxidoreductase subunit M	SBU_000879	802	57.6	
<i>nuoN</i>	NADH:quinone oxidoreductase subunit N	SBU_000880	994	74.9	

Expression as absolute read counts and as fragments per kilobase of transcript per million mapped reads (FPKM) is shown, as well as the corresponding protein abundance as emPAI index.

*Submitted as CoB-CoM heterodisulfide reductases.

Extended Data Table 6 | BLASTP search of proteins involved in butyrate oxidation. Best results according to the E-value are shown

Gene product	Locus tag	Blast annotation	Accession number	Organism	Coverage /Identity	E-value
Acyl-CoA dehydrogenase domain-containing protein	SBU_000172	acyl-CoA dehydrogenase	WP_028324559.1	<i>Desulfatirhabdium butyrativorans</i>	100/66	0.0
Acyl-CoA dehydrogenase	SBU_000399	acyl-CoA dehydrogenase	WP_028321791.1	<i>Desulfatiglans anilini</i>	99/51	6e-175
Acyl-CoA dehydrogenase domain-containing protein	SBU_000724	acyl-CoA dehydrogenase	WP_028319806.1	<i>Desulfatiglans anilini</i>	98/62	8e-175
Acyl-CoA dehydrogenase	SBU_001146	hypothetical protein	WP_036734863.1	<i>Peptococcaceae bacterium SCADC1_2_3</i>	97/58	0.0
Crotonase	SBU_000400	hypothetical protein	WP_029475295.1	<i>Dehalococcoidia bacterium SCGC AB-539-J10</i>	96/69	7e-128
3-Hydroxyacyl-CoA dehydrogenase	SBU_000288	3-hydroxybutyryl-CoA dehydrogenase Hbd	EMS78924.1	<i>Desulfotignum phosphitoxidans</i> DSM 13687	100/65	8e-134
3-Hydroxyacyl-CoA dehydrogenase	SBU_000843	3-hydroxy-2-methylbutyryl-CoA dehydrogenase	WP_007907297.1	<i>Ktedonobacter racemifer</i>	98/61	5e-98
Acetyl-CoA acetyltransferase	SBU_000329	acetyl-CoA acetyltransferase	WP_014407007.1	<i>Methanocella conradii</i>	99/67	0.0
Acetyl-CoA acetyltransferase	SBU_000402	conserved hypothetical protein, thiolase family	CBH39006.1	uncultured archaeon	100/70	0.0
Acetyl-CoA acetyltransferase	SBU_000404	conserved hypothetical protein, thiolase family	CBH39006.1	uncultured archaeon	100/57	6e-159
Acetyl-CoA acetyltransferase	SBU_001291	acetyl-CoA acetyltransferase	WP_010917519.1	<i>Thermoplasma volcanium</i>	98/43	4e-103

Extended Data Table 7 | Genes encoding Type IV pili and 10 most expressed cytochromes identified in the HotSeep-1 genome bin from the Butane50 culture

Gene	Feature	Best Blast Hit		Pfam hits	Transcriptome		
		Accession number	Coverage/Identity	Domain ID/ accession	Absolute reads	FPKM	
PilA	type IV pilus assembly protein	AMM41792.1	100/99	N_methyl_3/ PF13633.3	8	3	
PilA	pilus assembly protein PilA	AMM39924.1	100/88	N_methyl_3/ PF13633.3	1780	582	
Peptidase A24	peptidase A24	AMM42236.1	100/100	Dis_P_Dis/ PF06750; Peptidase_A24/ PF01478	31	9	
PilQ	pilus modification protein PilQ	AMM40410.1	100/99	AMIN/ PF11741; STN/ PF07660; Secretin_N/ PF03958; Secretin/PF00263	116	12	
PilP	type IV pilus assembly protein PilP	AMM40409.1	100/99	T2SSC/ PF11356; PilP/PF04351	8	4	
PilO	pilus assembly protein PilO	AMM40408.1	100/100	PilO/PF04350	7	2	
PilN	type IV pilus assembly protein PilN	AMM40407.1	100/98	PilN/PF05137	13	5	
PilM	pilus assembly protein PilM	AMM40406.1	100/99	Pil_2/PF11104	32	6	
PilW	putative pilus assembly protein PilW	AMM39834.1	100/99	N_methyl_2/PF13544	53	10	
PilY1	type IV pilus assembly protein	AMM41700.1	100/98	-	247	12	
PilY1	putative pilus assembly protein PilY	AMM41699.1	100/99	Neisseria_PilC/PF05567	91	16	
PilL	type IV pilus assembly protein pilL	AMM41701.1	100/93	-	3	2	
PilC	type IV pilus assembly protein PilC	AMM42043.1	97/99	T2SSF/PF00482 (x2)	50	9	
Cytochrome c type based on PfamA domain prediction	Feature	Best Blast Hit Accession number	Coverage/I dentity	Cellular localization (PSORTb) ¹	Heme groups	Transcriptome Absolute reads	FPKM
Cytochrom_CIII	class III cytochrome C	AMM42051.1	100/97	Periplasmic	4	505	238
Paired_CXXCH_1	doubled CXXCH cytochrome C	AMM40456.1	99/46	Extracellular	5	1244	232
Paired_CXXCH_1	doubled CXXCH cytochrome C	AMM40456.1	100/94	Unknown (CM,P,OM,E)	6	768	143
Paired_CXXCH_1	doubled CXXCH cytochrome C	AMM39976.1	100/94	Unknown (CM,P,OM,E)	7	471	94
Paired_CXXCH_1	cytochrome C	AMM40455.1	96/95	Extracellular	6	483	96
Cytochrom_CIII	class III cytochrome C	AMM41048.1	100/99	Periplasmic	4	69	31
Paired_CXXCH_1	cytochrome C	AMM40346.1	100/99	Cytoplasmic	10	135	28
Paired_CXXCH_1	cytochrome C	AMM40455.1	99/46	Periplasmic	7	105	21
Cytochrom_c3_2	cytochrome C	AMM40350.1	100/100	Unknown (CM,P,E)	12	72	17
Cytochrom_c3_2	cytochrome C	AMM40349.1	100/99	Periplasmic	12	53	14

¹For unknown cellular localization of cytochromes, potential locations are indicated according to the score value; CM, cytoplasmic membrane; P, periplasmic; OM, outer membrane; E, extracellular.

Tidal evolution of the Moon from a high-obliquity, high-angular-momentum Earth

Matija Ćuk¹, Douglas P. Hamilton², Simon J. Lock³ & Sarah T. Stewart⁴

In the giant-impact hypothesis for lunar origin, the Moon accreted from an equatorial circum-terrestrial disk; however, the current lunar orbital inclination of five degrees requires a subsequent dynamical process that is still unclear^{1–3}. In addition, the giant-impact theory has been challenged by the Moon's unexpectedly Earth-like isotopic composition^{4,5}. Here we show that tidal dissipation due to lunar obliquity was an important effect during the Moon's tidal evolution, and the lunar inclination in the past must have been very large, defying theoretical explanations. We present a tidal evolution model starting with the Moon in an equatorial orbit around an initially fast-spinning, high-obliquity Earth, which is a probable outcome of giant impacts. Using numerical modelling, we show that the solar perturbations on the Moon's orbit naturally induce a large lunar inclination and remove angular momentum from the Earth–Moon system. Our tidal evolution model supports recent high-angular-momentum, giant-impact scenarios to explain the Moon's isotopic composition^{6–8} and provides a new pathway to reach Earth's climatically favourable low obliquity.

The leading theory for lunar origin is the giant impact^{9,10}, which explains the Moon's large relative size and small iron core. Here we refer to the giant-impact theory in which the Earth–Moon post-impact angular momentum (AM) was the same as it is now (in agreement with classic lunar tidal evolution studies^{11,12}) as the 'canonical' theory. In the canonical giant-impact model¹³, a Mars-mass body obliquely impacts the proto-Earth near the escape velocity to generate a circum-terrestrial debris disk. The AM of the system is set by the impact, and the Moon accretes from the disk, which is predominantly (>60 wt%) composed of impactor material. However, Earth and the Moon share nearly identical isotope ratios for a wide range of elements, and this isotopic signature is distinct from all other extraterrestrial materials^{4,5}. Because isotopic variations arise from multiple processes⁴, the Moon must have formed from, or equilibrated with, Earth's mantle^{5,14}. Earth–Moon isotopic equilibration in the canonical model has been proposed by ref. 15, but has been questioned by other researchers¹⁶, who suggest that the large amount of mass exchange required to homogenize isotopes could lead to the collapse of the proto-lunar disk.

Ćuk and Stewart⁶ proposed a new variant of the giant-impact model that is based on an initially high-AM Earth–Moon system. In this model, a late erosive impact onto a fast-spinning proto-Earth produced a disk that was massive enough to form the Moon, and was composed primarily of material from Earth, potentially satisfying the isotopic observations. Canup⁷ presented a variation of a high-AM origin in which a slow collision between two similar-mass bodies produces a fast-spinning Earth and a disk with Earth-like composition that becomes the Moon. Subsequently, Lock *et al.*⁸ have argued that a range of high-energy, high-AM giant impacts generate a particular post-impact state where the Earth's mantle, atmosphere and disk are not dynamically isolated from each other, enabling widespread mixing and equilibration between the accreting Moon and Earth. After the

impact, these high-AM models require a plausible mechanism with which to remove AM to be consistent with the current Earth–Moon system.

Ćuk and Stewart⁶ originally proposed that the excess AM was lost during tidal evolution of the Moon via the evection resonance between Earth's orbital period and precession of the Moon's perigee⁶. More recently, Wisdom and Tian¹⁷ found that the evection near resonance can reduce the system's AM to the present value for a wider range of tidal parameters than explored by ref. 6. However, AM loss through the evection resonance is still confined to a subset of possible tidal parameters for Earth and the Moon, and the high-AM giant-impact scenario requires a robust mechanism for reproducing the present-day system. In this work, we propose a more plausible model of lunar tidal evolution that removes AM, but also may solve another major problem for lunar origin: the Moon's orbital inclination.

The Moon's orbit is currently inclined by about 5°, but studies of its tidal evolution^{11,12} have found that the inclination would have to have been at least 12° at formation, if the inclination was primordial. This is at odds with lunar formation from a flat disk in Earth's equatorial plane, which should produce a Moon with no inclination. Hypotheses that have been proposed to explain the lunar inclination include a sequence of Moon–Sun resonances¹, resonant interaction with the protolunar disk², and encounters with large planetesimals after the formation of the Moon³. However, past studies of lunar tidal history^{6,11,12} ignored the obliquity tides within the Moon, despite the Moon having very large 'forced' obliquity when it was between 30 and 40 Earth radii (R_E) away from Earth owing to the lunar spin axis undergoing the Cassini state transition^{18,19}. Chen and Nimmo²⁰ found that the lunar obliquity tides (driven by Earth's apparent north–south motion relative to the lunar figure (body)) greatly decrease the Moon's orbital inclination.

To quantify the effect of obliquity tides, we used a semi-analytical tidal model (see Methods subsection 'Damping of lunar inclination by obliquity tides'). Chen and Nimmo²⁰ considered tides within the lunar magma ocean that rely on excitation of Rossby waves²¹, but here we considered only the tidal response of the current, 'cold' Moon. If we additionally assume long-term average tidal dissipation within Earth and current lunar tidal properties, we find that the orbital inclination of the Moon must have been substantially higher before the Cassini state transition, possibly as high as 30° (Methods subsection 'Damping of lunar inclination by obliquity tides' and Extended Data Fig. 1). If we were to extrapolate this inclination back to the time of lunar formation close to Earth, the Moon must have formed with an orbit inclined at over 50° to the equator, which is clearly inconsistent with a giant-impact origin and suggests that the inclination was acquired after the Moon formed. In addition, the planetesimal encounter hypothesis³ has difficulty producing both the correct lunar eccentricity and inclination simultaneously when lunar inclination damping by lunar obliquity tides and lunar eccentricity excitation by Earth tides are taken into account

¹Carl Sagan Center, SETI Institute, 189 North Bernardo Avenue, Mountain View, California 94043, USA. ²Department of Astronomy, University of Maryland, Physical Sciences Complex, College Park, Maryland 20742, USA. ³Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA. ⁴Department of Earth and Planetary Sciences, University of California Davis, One Shields Avenue, Davis, California 95616, USA.

(Methods subsection ‘Excitation of lunar inclination by encounters with planetesimals’ and Extended Data Fig. 2).

We show that the tidal evolution of the Moon starting with a high-obliquity, high-AM Earth can reproduce the current lunar orbit, including the lunar inclination and the Earth–Moon system AM. For any perturbed orbit, there exists a Laplace plane around which the orbital plane of the perturbed orbit precesses. For close-in moons of oblate planets like Jupiter and Saturn, the Laplace plane is the equatorial plane of the planet, whereas for outer irregular satellites of these planets that are strongly perturbed by the Sun, the Laplace plane is their planet’s heliocentric orbital plane. For the Moon, the Laplace plane undergoes a transition during lunar tidal evolution when the Moon recedes from the inner region dominated by perturbations from Earth’s equatorial bulge to the outer region dominated by solar perturbations. In the inner region, the Laplace plane is close to Earth’s equator, and in the outer region, the Laplace plane is close to the plane of Earth’s heliocentric orbit (the ecliptic). At the transition between these two regimes, the Laplace plane is intermediate between the equator and the ecliptic. The distance at which the Laplace plane transition occurs is approximately²²

$$r_L = \left(2J_2 \frac{M_E}{M_S} R_E^2 a_E^3 \right)^{1/5} \quad (1)$$

where J_2 is the oblateness moment of Earth, M_E and M_S are the mass of Earth and the mass of the Sun, and a_E is Earth’s semimajor axis. For a planet like Earth that is in hydrostatic equilibrium, J_2 depends on the rotation rate. Therefore, the critical distance for the Laplace plane transition has been moving inward over the course of lunar tidal evolution as Earth’s rotation slows down and Earth becomes more spherical. For an Earth–Moon system with 100%–180% of the present AM, the Laplace plane transition happens at 16–22 R_E .

For small and moderate obliquities of Earth (that is, angles between the Equator and the ecliptic), the Laplace plane transition is smooth and does not produce any excitation of lunar eccentricity or inclination. However, the Laplace plane transition causes orbital instabilities for obliquities above 68.9° (ref. 23). Satellites on circular orbits around high-obliquity planets migrating through the Laplace plane transition can acquire substantial eccentricities and inclinations. This excitation is driven by solar secular perturbations that operate at high inclinations (“Kozai resonance”²⁴). For high planetary obliquities, satellites close to the Laplace plane transition with low ‘free’ inclinations (to the local Laplace plane) still experience solar Kozai perturbations, because their orbits have large instantaneous inclinations relative to the ecliptic²⁵. A related mechanism produces complex dynamics previously found by ref. 26 for the tidal evolution of hypothetical high-obliquity Earth-like planets with large moons. Atobe and Ida²⁶ also found that the mass of a Moon-sized satellite has a large effect on the dynamics of the system, enabling stagnation or even reversal of tidal evolution and large-scale AM loss from the system; however, they used an averaged model that did not track eccentricity and could not capture the full dynamics of the Laplace plane transition.

To study the tidal evolution of the Moon from a high-obliquity Earth followed by inclination damping at the Cassini state transition, we wrote a specialized numerical integrator, R-SISTEM, which resolves lunar rotation and therefore fully models lunar obliquity tides (Methods subsection ‘Numerical methods’). Figure 1 (see also Supplementary Video 1) shows the early tidal evolution for two simulations that assume that the Earth–Moon system initially had 1.8 times its current AM, as proposed by ref. 6, but with an initial obliquity to the ecliptic of 70°. Solar perturbations induce sizeable lunar eccentricity when the Moon reaches the Laplace plane transition at about 17 R_E , triggering strong eccentricity-damping satellite tides that shrink the semimajor axis and approximately balance the outward push of Earth tides (see the first 16 Myr of Fig. 1). Because eccentric orbits have less AM than circular ones, AM is removed from the lunar orbit and transferred

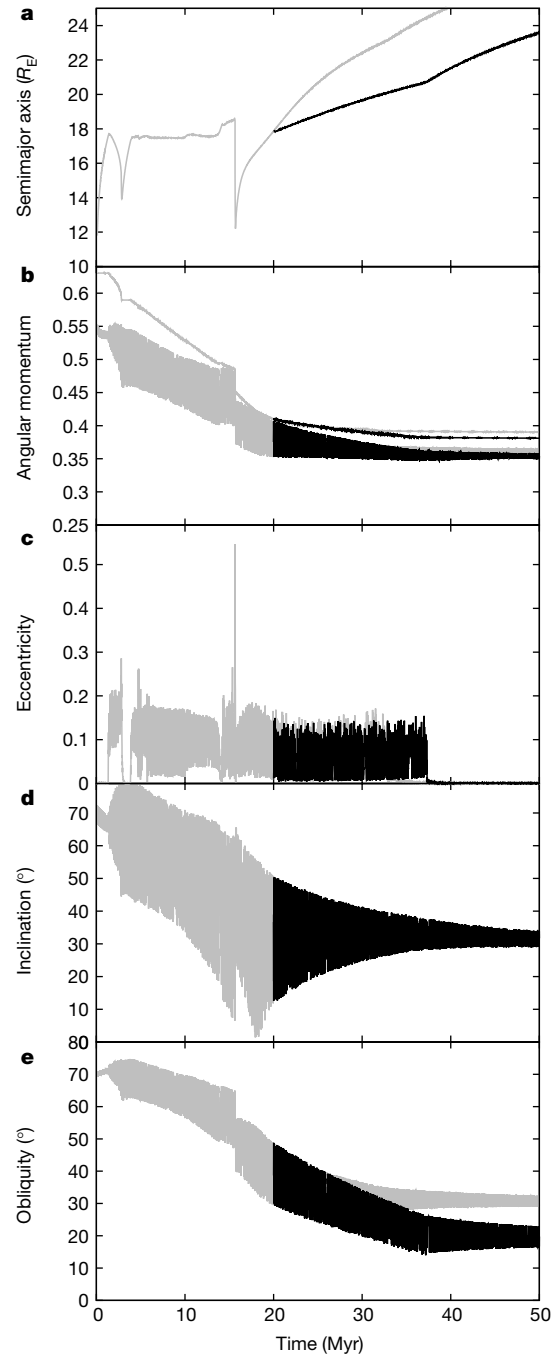


Figure 1 | Numerical simulation of the Moon’s early tidal evolution from Earth with initial obliquity of 70° and spin period of 2.5 h.

a, Semimajor axis. **c**, Eccentricity. **d**, Inclination to the ecliptic of the lunar orbit. **e**, Earth’s obliquity to the ecliptic and the AM of the system (**b**) in

units of $\alpha_E \sqrt{GM_E^3 R_E}$, where M_E , R_E and $\alpha_E = 0.33$ are the mass, radius and the scaled moment of inertia of Earth, respectively, and G is the gravitational constant. In these units, the present AM is 0.35. The grey lines plot a simulation in which tidal properties of Earth and the Moon were $k_{2,E}/Q_E = k_{2,M}/Q_M = 0.01$ throughout (see Methods subsection ‘Damping of lunar inclination by obliquity tides’ for definitions). The black line shows a simulation branching at 20 million years (Myr) after the giant impact by changing $k_{2,E}/Q_E$ to 0.005. In the AM plot (**b**), the thin lines plot a scalar sum of spin and orbital AM, while the lower (thick) band includes only the component of the lunar orbital AM vector that is perpendicular to the ecliptic.

to Earth’s heliocentric orbit; Earth tides in turn transfer AM from Earth’s spin to the lunar orbit, whereas satellite tides do not change the AM of the Earth–Moon system. During this prolonged stalling of lunar

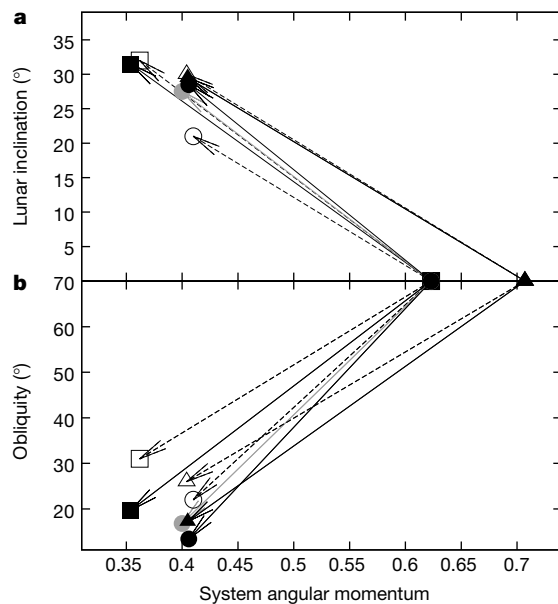


Figure 2 | Lunar inclination and Earth's obliquity for various simulations. The initial and final lunar inclination (a) and Earth's obliquity (b) are plotted against corresponding values of the Earth–Moon system AM in different simulations. AM is plotted in units of $\alpha_E \sqrt{GM_E^3 R_E}$, and all simulations of the Laplace plane transition included here started with Earth's obliquity of 70° . The squares correspond to simulations plotted in Fig. 1: the open square plots a simulation with $Q_E/k_{2,E} = Q_M/k_{2,M} = 100$ and the filled square plots the branch for which $Q_E/k_{2,E} = 200$ after 20 Myr. Circles plot simulations with $Q_E/k_{2,E} = 200$ throughout (shown in Extended Data Fig. 5), with white, grey and black circles corresponding to cases with $Q_M/k_{2,M} = 200, 100$ and 50 , respectively. Simulations with higher initial AM (2.25 h initial spin) (Extended Data Fig. 6) are plotted with triangles: the open triangle plots a simulation with $Q_E/k_{2,E} = Q_M/k_{2,M} = 100$ and the filled triangle plots the branch for which $Q_E/k_{2,E} = 200$ after 30 Myr. All simulations saw a substantial loss of AM, excitation of lunar inclination to large values, and a large reduction of Earth's obliquity. We conclude that a smaller Q_E leads to a larger AM loss, while a greater Q_E later in the Laplace plane transition leads to smaller final obliquity for Earth.

tidal evolution, the Moon acquires large inclination (over 30°), while the obliquity of Earth decreases.

In the later part of this complex period of the lunar orbital history, lunar eccentricity is excited by secular near-resonances between lunar inclination and eccentricity and Earth's oblateness (see Methods subsection 'Laplace plane transition', and Extended Data Figs 3 and 4). Depending on the exact tidal parameters used for Earth and the Moon, Earth's obliquity can reach that required to match the present value ($<20^\circ$)²⁷, while the AM of the system also matches the present value (0.35 in units of $\alpha_E \sqrt{GM_E^3 R_E}$; see Fig. 1 legend for definitions). Figure 2 presents the results of the Laplace plane transition simulations using different tidal parameters for Earth and the Moon and a different initial AM of Earth (Extended Data Figs 5 and 6). Large AM loss, high lunar inclination and low terrestrial obliquity is a frequent outcome.

In Fig. 3, we explore the early part of the Laplace plane transition for different initial obliquities of Earth (with the same total AM). Cases with initial obliquities of 65° and 75° are similar to Fig. 1, with larger obliquity leading to larger AM loss in the early stages of the Laplace plane transition. The simulation with an initial obliquity of 80° experiences stalling and reversal of tidal evolution, with the Moon eventually falling back onto Earth. This evolution agrees with prior results²⁶, and we expect this outcome for all obliquities larger than 80° . We also found that the Moon evolving from Earth with an initial obliquity of 60° does not experience any instability at the Laplace plane transition. As the orientation of terrestrial planets' spin axes is probably determined by

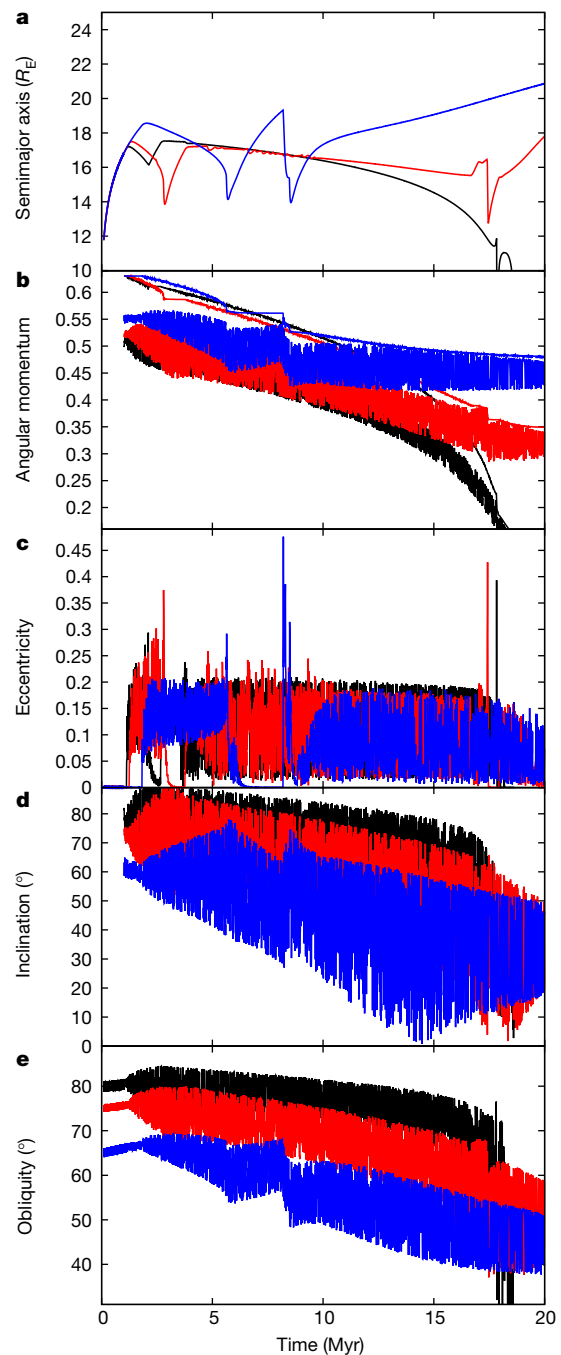


Figure 3 | Similar to Fig. 1, but with different initial obliquities for Earth. Obliquities are 80° (black line), 75° (red line) and 65° (blue line). $Q_E/k_{2,E} = Q_M/k_{2,M} = 100$ throughout the simulations. Unlike the simulations shown in Fig. 1, these simulations were followed for only 20 Myr. The 80° simulation leads to the Moon falling back on Earth. We note a trend that the AM loss (b) is larger for higher initial obliquities of Earth.

giant impacts, their poles should be randomly distributed on a sphere. About a third of all Earth-like planets should have obliquities within 60° – 80° or 100° – 120° , and, if they have large moons, would lose AM and obliquity (without losing the moon) at the Laplace plane transition. Furthermore, Kokubo and Genda²⁸ found that planets with large AM are a frequent model outcome of terrestrial planet formation.

Once the Moon has passed through the Laplace plane transition, it continues to recede from Earth and lunar rotation passes through the Cassini state transition¹⁸. Using R-SISTEM, we find that the Moon probably spent some time in a non-synchronous rotation state when

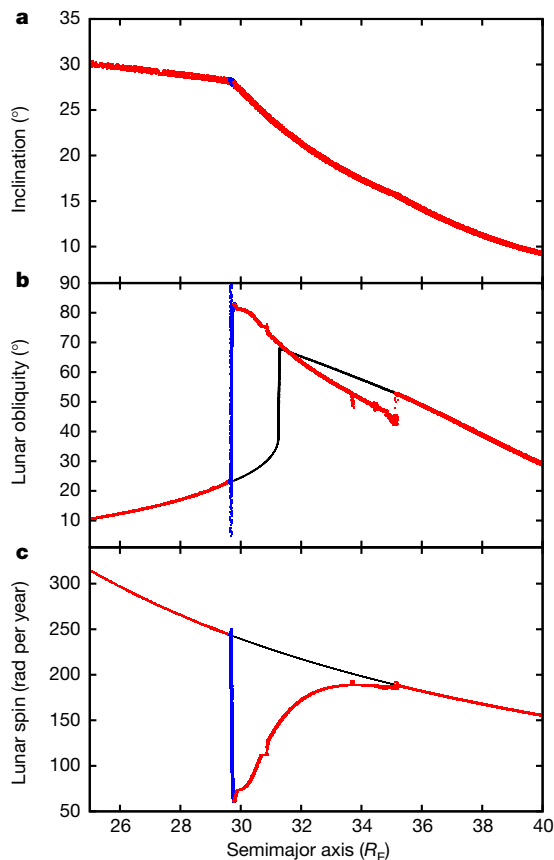


Figure 4 | Numerical integration of the later phase of lunar tidal evolution, assuming a lunar inclination of 30° at $25R_E$ and the current shape of the Moon. The panels plot lunar inclination to the ecliptic (a), lunar obliquity with respect to its orbit (b) and lunar spin rate (c) against the Moon's semimajor axis. The red points plot the segments of the tidal evolution that were artificially accelerated, while the blue points plot the intervals integrated with nominal parameters. We used $k_{2,E} = 0.3$, $Q_E = 35$, $k_{2,M} = 0.024$ and $Q_{M,0} = 60$ (for the numerical implementation of lunar tides, see Methods subsection 'Numerical methods'), with the Love numbers enhanced by 100 times before the event at $29.7R_E$ and by 10 times after that event. Black lines plot the Cassini state obliquity (b) and the synchronous rotation rate (c) expected for each instantaneous semimajor axis and inclination. The Moon is in non-synchronous rotation from $29.7R_E$ to $35R_E$.

close to the Cassini state transition (Fig. 4, Extended Data Figs 7 and 8) and that transitions between the rotation states can be triggered by various resonances or impacts (Methods subsection 'Cassini state transition', Extended Data Figs 9 and 10, and Supplementary Video 2). Regardless of the nature of the lunar rotation state, the Moon's obliquity is very high during the Cassini state transition and immediately following it, leading to damping of lunar inclination. Although the implementation of high-obliquity satellite tides in a fully numerical integrator is challenging (Methods subsection 'Numerical methods'), we find that the lunar inclination damps from 30° (obtained during the Laplace plane transition) to its present value if we assume the long-term average tidal properties for Earth and a relatively non-dissipative, solid Moon (Fig. 4).

The rotational dynamics of the Moon is strongly dependent on the Moon's global shape. The early Moon probably had little strength, and its shape was in equilibrium with tidal forces²⁹. In Fig. 1, we modelled the Moon as a rigid body for numerical tractability, but we periodically reset its figure to match an equilibrium shape for that distance from Earth, assuming synchronous rotation³⁰. This assumption of a hydrostatic-like shape results in low obliquities in Cassini state 1 when the Moon is close to Earth¹⁸. Since the current triaxial shape of the

Moon matches the order of magnitude of tidal deformation expected at $23\text{--}26R_E$ (ref. 29), we assumed that the Moon is rigid and has the present-day principal moments beyond $25R_E$. Our orbital history, shown in Fig. 1, may be consistent with the proposal that the current lunar shape froze at a distance of $15\text{--}17R_E$ from Earth on an orbit with $e \approx 0.2$ (ref. 30).

Our high-obliquity model is at present the only model we are aware of to explain the origin of large past lunar inclination, which was subsequently reduced by strong obliquity tides at the Cassini state transition²⁰. A high-obliquity early Earth offers an AM loss mechanism more robust than the evection resonance^{6,17}. Therefore, our results support high-AM giant-impact scenarios for lunar origin^{6–8}. An initially high-obliquity Earth is consistent with the expectation of random spin-axis orientations for terrestrial planets after giant impacts, and the dynamics discussed here naturally reduces Earth's obliquity to values that are low to moderate. This mechanism also provides a novel way in which initially highly tilted terrestrial exoplanets can acquire low obliquities and potentially stable climates.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 April; accepted 9 September 2016.

Published online 31 October 2016.

1. Touma, J. & Wisdom, J. Resonances in the early evolution of the Earth–Moon system. *Astron. J.* **115**, 1653–1663 (1998).
2. Ward, W. R. & Canup, R. M. Origin of the Moon's orbital inclination from resonant disk interactions. *Nature* **403**, 741–743 (2000).
3. Pahlevan, K. & Morbidelli, A. Collisionless encounters and the origin of the lunar inclination. *Nature* **527**, 492–494 (2015).
4. Burkhardt, C. *Isotopic Composition of the Moon and the Lunar Isotopic Crisis* 1–13 (Springer, 2015).
5. Young, E. D. *et al.* Oxygen isotopic evidence for vigorous mixing during the Moon-forming giant impact. *Science* **351**, 493–496 (2016).
6. Cuk, M. & Stewart, S. T. Making the Moon from a fast-spinning Earth: a giant impact followed by resonant despinning. *Science* **338**, 1047–1052 (2012).
7. Canup, R. M. Forming a Moon with an Earth-like composition via a giant impact. *Science* **338**, 1052–1055 (2012).
8. Lock, S. J. *et al.* A new model for lunar origin: equilibration with Earth beyond the hot spin stability limit. *Lunar Planet. Sci. Conf.* **47**, 2881 (2016).
9. Hartmann, W. K. & Davis, D. R. Satellite-sized planetesimals and lunar origin. *Icarus* **24**, 504–515 (1975).
10. Cameron, A. G. W. & Ward, W. R. The origin of the moon. *Lunar Planet. Sci. Conf.* **7**, 120 (1976).
11. Goldreich, P. History of lunar orbit. *Rev. Geophys.* **4**, 411–439 (1966).
12. Touma, J. & Wisdom, J. Evolution of the Earth–Moon system. *Astron. J.* **108**, 1943–1961 (1994).
13. Canup, R. M. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412**, 708–712 (2001).
14. Meier, M. M. M. Earth's titanium twin. *Nat. Geosci.* **5**, 240–241 (2012).
15. Pahlevan, K. & Stevenson, D. J. Equilibration in the aftermath of the lunar-forming giant impact. *Earth Planet. Sci. Lett.* **262**, 438–449 (2007).
16. Melosh, H. J. New approaches to the Moon's isotopic crisis. *Phil. Trans. R. Soc. Lond. A* **372**, <http://dx.doi.org/10.1098/rsta.2013.0168> (2014).
17. Wisdom, J. & Tian, Z. Early evolution of the Earth–Moon system with a fast-spinning Earth. *Icarus* **256**, 138–146 (2015).
18. Ward, W. R. Past orientation of the lunar spin axis. *Science* **189**, 377–379 (1975).
19. Chyba, C. F., Jankowski, D. G. & Nicholson, P. D. Tidal evolution in the Neptune–Triton system. *Astron. Astrophys.* **219**, L23–L26 (1989).
20. Chen, E. M. A. & Nimmo, F. Tidal dissipation in the early lunar magma ocean and its effect on the evolution of the Earth–Moon system. *Icarus* **275**, 132–142 (2016).
21. Tyler, R. H. Strong ocean tidal flow and heating on moons of the outer planets. *Nature* **456**, 770–772 (2008).
22. Nicholson, P. D., Cuk, M., Sheppard, S. S., Nesvorný, D. & Johnson, T. V. *In Irregular Satellites of the Giant Planets* 411–424 (Univ. Arizona Press, 2008).
23. Tremaine, S., Touma, J. & Namouni, F. Satellite dynamics on the Laplace surface. *Astron. J.* **137**, 3706–3717 (2009).
24. Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.* **67**, 591 (1962).
25. Tamayo, D., Burns, J. A., Hamilton, D. P. & Nicholson, P. D. Dynamical instabilities in high-obliquity systems. *Astron. J.* **145**, 54 (2013).
26. Atobe, K. & Ida, S. Obliquity evolution of extrasolar terrestrial planets. *Icarus* **188**, 1–17 (2007).

27. Rubincam, D. P. Tidal friction in the Earth–Moon system and Laplace planes: Darwin redux. *Icarus* **266**, 24–43 (2016).
28. Kokubo, E. & Genda, H. Formation of terrestrial planets from protoplanets under a realistic accretion condition. *Astrophys. J. Lett.* **714**, 21–25 (2010).
29. Garrick-Bethell, I., Wisdom, J. & Zuber, M. T. Evidence for a past high-eccentricity lunar orbit. *Science* **313**, 652–655 (2006).
30. Keane, J. T. & Matsuyama, I. Evidence for lunar true polar wander and a past low-eccentricity, synchronous lunar orbit. *Geophys. Res. Lett.* **41**, 6610–6619 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by NASA's Emerging Worlds programme, award NNX15AH65G.

Author Contributions M.Ć. designed the study, wrote the software, analysed the data and wrote the paper. Co-authors contributed ancillary calculations, discussed the results and wider implications of the work, and helped edit and improve the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Ć. (mcuk@seti.org).

Reviewer Information *Nature* thanks D. Stevenson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Damping of lunar inclination by obliquity tides. To quantify the effect of obliquity tides on the lunar orbit, we constructed a semi-analytical method for modelling the evolution of lunar orbit under the influence of Earth and Moon tides. This model assumes that the Moon is in synchronous rotation and the relevant Cassini state, unlike the fully numerical integrator R-SISTEM that we use elsewhere in this work, which fully resolves lunar rotation. The goal of this model is (1) simply to demonstrate the importance of lunar obliquity tides during the Cassini state transition, and (2) to allow for more efficient integration of the lunar orbit beyond $40R_E$, when the Moon was probably in Cassini state 2, and its obliquity, inclination and eccentricity were moderate or small.

The semi-analytical model is described by a system of equations^{19,31,32}:

$$\left(\frac{da}{dt}\right)_E = \frac{3k_{2,E}M_M}{Q_E M_E} \left(\frac{R_E}{a}\right)^5 an \quad (1)$$

$$\left(\frac{de}{dt}\right)_E = \frac{19}{8} \frac{e}{a} \left(\frac{da}{dt}\right)_E \quad (2)$$

$$\left(\frac{di}{dt}\right)_E = -\frac{1}{4} \frac{\sin i}{a} \left(\frac{da}{dt}\right)_E \quad (3)$$

$$\left(\frac{de}{dt}\right)_M = -e \frac{21k_{2,M}M_E}{2Q_M M_M} \left(\frac{R_M}{a}\right)^5 n \quad (4)$$

$$\left(\frac{di}{dt}\right)_M = \frac{-\sin^2 \theta}{\tan i} \frac{3k_{2,M}M_E}{2Q_M M_M} \left(\frac{R_M}{a}\right)^5 n \quad (5)$$

$$\left(\frac{da}{dt}\right)_M = 2ae \left(\frac{de}{dt}\right)_M + 2a \tan i \left(\frac{di}{dt}\right)_M \quad (6)$$

where a , e , i and n are the semimajor axis, eccentricity, inclination and mean motion of the lunar orbit, respectively; t is time; M , R , k_2 and Q are mass, radius, Love number and tidal quality factor, while the subscripts E and M refer to Earth and Moon, respectively (subscripts after derivatives indicate whether the effect is due to Earth or lunar tides). Lunar obliquity to the orbit, θ , is calculated as the solution to the equation^{18,33}

$$\frac{2}{3} n \left(\frac{C-A}{C} \right) \sin \theta \cos \theta + \frac{3}{8} n \left(\frac{B-A}{C} \right) \sin \theta (1 - \cos \theta) + \dot{\Omega} \sin(\theta - i) = 0 \quad (7)$$

where A , B and C are the Moon's principal moments of inertia, and $\dot{\Omega}$ is the rate of precession of the lunar node, which we estimate as³⁴

$$\dot{\Omega} = \frac{3}{4} \frac{n_E^2}{n} \cos i \quad (8)$$

where n_E is Earth's heliocentric mean motion. This expression for $\dot{\Omega}$ assumes precession dominated by solar perturbations and is invalid close to Earth, where the influence of its equatorial bulge is important. Our numerical implementation is a simple mapping that solves for lunar obliquity, calculates tidal derivatives and then advances the orbital elements (we used a 1-Myr timestep, appropriate for distances beyond approximately $25R_E$).

Extended Data Figure 1 plots three orbital histories calculated using the above approach that result in the Moon moving from $25R_E$ to $60R_E$ over 4.5 billion years, and having a final inclination of about 5° (the Moon's eccentricity was about 0.01 in these calculations). Extended Data Figure 1 clearly shows that the past studies of lunar tidal evolution that neglected lunar tides^{11,12} greatly underestimate the past inclination of the Moon. Even assuming a tidal quality factor of $Q = 100$ (which is larger than the present value of 38) for the Moon during the Cassini state transition, the inclination had to be over 17° at $25R_E$, which would in turn require an inclination of approximately 30° when the Moon was close to Earth, much greater than the often-cited 12° (ref. 12). If we adopt the current tidal parameters of the Moon ($Q = 38$), the inclination at $25R_E$ was over 30° , seemingly at odds with the formation of the Moon in an equatorial disk.

Although our simplified model uses several approximations, we can argue that the values for the past lunar inclination that we calculate are underestimates, for at least three reasons. First, equation (5) assumes low obliquities, and at obliquities close to 90° , the leading coefficient is $15/8$ rather than $3/2$ (ref. 19), leading to a more rapid damping of inclination. Second, we assumed the current Love

number for the Moon ($k_{2,M} = 0.024$), which implies that the Moon was as rigid at the Cassini state transition as it is now; however, because the Moon was only a few hundred million years old and much warmer in the interior at the time of the transition, its Love number would be higher than it is now, again leading to more rapid inclination damping. Third, we assumed constant tidal properties for Earth throughout the calculation, which is in conflict with the fact that the current rate of the Moon's tidal recession is about three times higher than the long-term average³⁵. An increase in the tidal dissipation within Earth's oceans over time, as suggested by modelling^{36,37} would mean that the Moon spent more time at the Cassini state transition than we calculated, allowing more damping of inclination. These factors are all independent of the fact that the Moon may not have been in synchronous rotation during the Cassini state transition, as discussed in this work.

Excitation of inclination by planetesimal fly-bys. Recently, Pahlevan and Morbidelli³ proposed that the lunar inclination was produced by encounters between the Earth–Moon system and leftover planetesimals. While innovative, this model has several outstanding issues which caused the authors substantially to overestimate the effectiveness of planetesimal encounters in raising the lunar inclination. They³ do not include lunar obliquity tides explicitly in their model, but assume that very high tidal dissipation rates within Earth moved the Moon to $40R_E$ within a few tens of millions of years, allowing most of the planetesimal encounters to occur after the Cassini state transition, keeping the newly acquired inclination safe from further damping. This timeline works if $k_{2,E}/Q_E = 0.1$ for the early Earth (which we find surprisingly high for a planetary body), but not for Earth's long-term average $k_{2,E}/Q_E = 0.01$, which would put the Moon at the Cassini state transition during the epoch of planetesimal encounters. However, given the uncertainties in the tidal properties of early Earth, here we will concentrate instead on the orbital mechanics aspects of the Pahlevan and Morbidelli³ model.

Extended Data Figure 2 shows the tidal evolution of the Moon using our semi-analytical model (Methods subsection 'Damping of lunar inclination by obliquity tides'), starting with the end-state of the simulation featured in figure 1 of ref. 3 ($a = 47R_E$, $i = 5.8^\circ$). Eccentricity was not specified in ref. 3, but it was stated that the encounter simulations typically result in $e = 2\sin i$, so we have used $e = 0.2$. We find that there is, starting from these initial conditions, no combination of tidal parameters for Earth and the Moon that can result in the correct eccentricity and inclination for the Moon at the present epoch. If we assume a larger excitation of lunar inclination, the correspondingly larger lunar eccentricity (given the $e = 2\sin i$ condition) counteracts Earth tides and slow down lunar recession. Stronger Earth tides lead to faster outward evolution but also produce a net increase of lunar eccentricity (owing to equation (2)), potentially leading to a reversal of the Moon's orbital evolution through lunar eccentricity tides. A slowdown or reversal in lunar tidal recession would preclude the Moon from reaching its present distance in 4.5 billion years. Therefore we conclude that for planetesimal encounters to be able to explain the lunar inclination, the encounters must excite inclination without substantially exciting eccentricity (that is, leave the Moon with a lower eccentricity than that reported by ref. 3). Given the stochastic nature of the process, a small number of encounter outcomes will have a low eccentricity and high inclination. However, while possible, such scenarios are statistically unlikely and therefore not compelling.

Mechanisms other than tides have the potential to alter lunar eccentricity. It has been suggested that the current lunar eccentricity of 0.055 is the product of a resonance between the Moon and Jupiter that happens when the Moon is at about $53R_E$ (ref. 38). This resonance arises because the rate of precession of the lunar longitude of pericentre is commensurate with the mean motion of Jupiter ('jovian evection'). The rate of precession is (ignoring lunar inclination, which is not affected by the resonance)³⁴:

$$\dot{\varpi} = \frac{3}{4} \frac{n_E^2}{n} \frac{\sqrt{1-e^2}}{(1-e^2)^{3/2}} + \frac{225}{32} \frac{n_E^3}{n^2} \frac{1-e^2}{(1-e^2)^3} + O(n_E^3/n^3) \quad (9)$$

where e_E is the eccentricity of Earth's orbit. Because e_E varies rapidly on 10^5 -year timescales owing to Milanković cycles³⁹, capture is impossible and the Moon crosses the resonance many times in both directions. These numerous crossings result in a random walk in lunar eccentricity, within a band in a and e associated with the resonance³⁸. Since this band moves to higher e for higher a , it is possible to reach arbitrary large eccentricities through this random walk; however, it is likely that the Moon will sooner or later reach the outer boundary of the resonant band (defined by the resonance location for $e_E = 0$), and exit this chaotic region with finite eccentricity, as shown in ref. 38.

However, it is also possible that through the random walk the lunar orbit will evolve to lower eccentricities. This reduction of eccentricity cannot be arbitrarily effective: it must be smaller than the thickness of the band in e for the semimajor

axis at which the Moon enters the resonant band. Keeping ϖ and a constant, the thickness of the band in e is determined by the variation in e_E :

$$(1 - e_{E, \max}^2)^{3/2} = \frac{\sqrt{1 - e_{\max}^2}}{\sqrt{1 - e_{\min}^2}} \quad (10)$$

where $e_{E, \max}$ is the maximum eccentricity of Earth during Milanković cycles (the minimum is zero) and e_{\max} and e_{\min} are the boundaries of the band at the semimajor axis at which the Moon encounters the resonant band. Entering the band with e_{\max} , the Moon cannot acquire an eccentricity lower than e_{\min} . So if we square equation (10) and keep only the quadratic terms in eccentricities, we get $e_{\max}^2 - e_{\min}^2 = 3e_{E, \max}^2$ or, assuming that $\Delta e = e_{\max} - e_{\min} \ll e_{\max}$

$$\Delta e = \frac{3e_{E, \max}^2}{2e_{\max}} \quad (11)$$

Therefore, for $e_{E, \max} = 0.06$ (ref. 39), a starting eccentricity of $e = 0.2$ cannot be reduced by the jovian evection by more than about 0.03, so the jovian evection does not substantially change the implications of Extended Data Fig. 2, namely that the lunar eccentricity will probably remain high if it was substantially excited by planetesimal encounters when the Moon was already beyond $40R_E$.

Numerical methods. Our dynamics code R-SISTEM (Rotation-Symplectic Integrator with Solar Tides in the Earth–Moon system) directly integrates both the orbital and rotational motion of the Moon. The Moon is treated as rigid and tri-axial, and experiences tidal accelerations, the effects of Earth's equatorial bulge and solar perturbations. The orbital part of the integrator is a symplectic mixed-variable integrator based on the principles of ref. 40, with the specific implementation taken from ref. 41. The integrator assumes the Moon is on a Keplerian orbit, with all other forces (including solar perturbations) inserted as periodic 'kicks' in Cartesian coordinates. The orbital part of R-SISTEM overlaps substantially with the more general-purpose satellite dynamics code SIMPL⁴², which has been extensively tested. Like SIMPL, R-SISTEM can include other planets (or other satellites of Earth) as perturbers, but neither were present in any of the integrations included in this paper. With no other planets, Earth's orbit was essentially Keplerian, as R-SISTEM ignores back reaction from tides and Moon–Sun interactions on Earth's heliocentric orbit.

Integration of the Moon's rotation was based on the Lie–Poisson approach of ref. 43, with the Moon treated as a tri-axial rigid body torqued by Earth and the Sun. This enables direct modelling of the Moon's axial precession and Cassini states, and allows for capture of lunar rotation into spin–orbit resonances, such as synchronous rotation. One issue we had to deal with is suppression of tumbling (that is, Chandler wobble), which can be triggered by some spin–orbit interactions. We adopted an approach similar to that of ref. 44, where a torque perpendicular to the AM vector acts on the AM vector (in the rotating reference frame) to push the AM vector towards the z -axis (the axis of the largest principal moment of inertia). The intensity of this torque is adjusted to match the wobble damping timescales predicted by ref. 45. If the AM vector is closer to the long axis of the body (that is, that associated with the minimum moment of inertia), a similar torque pushes the AM vector away from the long axis. These torques have reversed signs in opposite hemispheres, so that the wobble damping is the same regardless of the sense or rotation (in the body-fixed reference frame).

Although the rotation of Earth is not resolved in our model, it tracks the changes in Earth's rotation period caused by the tides. The oblateness of Earth is adjusted at the end of each timestep, depending on the new value of the spin rate. We assumed that $J_2 \approx \omega_E^2$ holds for all spin periods (where J_2 is the standard oblateness moment and ω_E is Earth's spin rate), which may be inaccurate if Earth was very non-spherical owing to fast spin. Earth's axis is made to precess as a result of instantaneous torques by the Moon and the Sun on the equatorial bulge (that is, we assume that Earth is purely oblate and in principal axis rotation).

The most important part of the numerical approach is the tides. Ćuk and Stewart⁶ approximated lunar tides as a radial force counteracting the radial motion of the Moon. Although this implementation could match eccentricity damping (assuming that the Moon is in synchronous rotation), it is not useful in modelling obliquity tides (because a radial force is always in the plane of the orbit). We still include a radial lunar tide, but it accounts only for the damping expected from the actual, physical radial tide (as opposed to libration and obliquity tides). The expression for the radial force used in the integrator is:

$$F_r = -\dot{r} \frac{9k_{2,M}M_E}{Q_M M_M} \sqrt{\frac{GM_E}{r^3}} \left(\frac{R_M}{r} \right)^5 \quad (12)$$

where r and \dot{r} are the Moon's geocentric distance and radial velocity. Using the expression for \dot{r} (ref. 39)

$$\dot{r} = \frac{na}{\sqrt{1 - e^2}} e \sin f \quad (13)$$

where f is the Moon's true anomaly, and the expression for the eccentricity damping by generalized accelerations is^{39,46}

$$\frac{de}{dt} = \sqrt{\frac{a(1 - e^2)}{GM_E}} [F_r \sin f + F_t (\cos f + \cos E)] \quad (14)$$

where F_t is a tangential acceleration (not present here) and E is the eccentric anomaly. We can then write

$$\frac{de}{dt} = -e \frac{9k_{2,M}M_E}{Q_M M_M} na^{3/2} R_M^5 r^{-13/2} \sin^2 f \quad (15)$$

To the lowest order in eccentricity, this expression averages over an orbital period to

$$\frac{de}{dt} = -e \frac{9k_{2,M}M_E}{2Q_M M_M} \left(\frac{R_M}{a} \right)^5 n \quad (16)$$

This is the analytical expression for the eccentricity damping expected from radial tides for a moon in synchronous rotation³⁹. Although our radial force approximates the predicted damping of eccentricity to the lowest order, our approach is fundamentally different from that taken in analytical derivation. Although ref. 39 shows that energy loss is expected during a monthly cycle of tidal stresses, our integrator 'does not know' that the cycle repeats monthly. All of the forces acting in our code are based only on the current positions and velocities (linear or angular) of the Sun, Earth and Moon. Despite these differences in approach, our force form has a clear physical meaning: energy is lost when a somewhat inelastic Moon moves radially in Earth's tidal field, as the mechanical energy going into elastic deformation turns into heat. Our force is strictly radial, so it does not affect the AM of the lunar orbit, in agreement with previous treatments of satellite tides³⁹.

We have also applied the principle that only the instantaneous quantities can be used to calculate tidal forces due to librational and obliquity tides. They are not treated differently, but are combined in the same 'kick', as they both arise from the Moon's orientation moving relative to the Earth–Moon line. An equal and opposite torque is applied to the Moon's spin, enabling despinning and damping into the synchronous rotation. Our approach is based on analogy with the general case of tides on a non-synchronous body; if we assume that the Moon is rotating much faster than it orbits the Earth, the acceleration experienced by the Moon from tides raised by Earth on the Moon would be³⁹

$$F_t = \text{sign}(\dot{\phi} - n) \frac{3k_{2,M}M_E}{2Q_M M_M} \left(\frac{R_M}{a} \right)^5 an^2 \quad (17)$$

where $\dot{\phi}$ is the lunar spin rate. We may be tempted to apply this equation to the synchronous case, and assume that the lag angle is independent of synodic frequency (equivalent to having constant Q) so that the angular acceleration simply changes sign when the rotation is slower than orbital motion³⁹. We can estimate the effect of such tidal acceleration on eccentricity (with obliquity set to zero), using equation (14):

$$\frac{de}{dt} = (\dot{\phi} - \dot{f}) \frac{3k_{2,M}M_E}{2Q_M M_M} \left(\frac{R_M}{a} \right)^5 n \sqrt{1 - e^2} (\cos f + \cos E) \quad (18)$$

For small eccentricities and synchronous rotation, we can assume that $\dot{\phi} - \dot{f} = -2\text{encos}M$, (M being the mean anomaly) so we need to integrate the above expression over $M = [-\pi/2, \pi/2]$ with a negative sign, and over $M = [\pi/2, 3\pi/2]$ with a positive sign. That would give us an expression for eccentricity damping that does not depend on eccentricity, which is certainly nonphysical. This is because strict constant- Q tides behave like a step function at the exact synchronous motion, without any sensitivity to the amount of deviation from synchronicity.

However, if we assume that the tidal Q does depend on frequency, according to $Q^{-1} = Q_0^{-1} 2(\dot{\phi} - \dot{f})/n$, and keep only the lowest terms in eccentricity, we get

$$\frac{de}{dt} = -6e \frac{k_{2,M}M_E}{Q_M M_M} \left(\frac{R_M}{a} \right)^5 n \quad (19)$$

which is the correct form for the eccentricity damping by the libration tide within a synchronous satellite³⁹. Therefore, in order to have a unified treatment of tides

in synchronous (or near-synchronous) and non-synchronous cases, we need to assume that the lag angle is proportional to the angular velocity of libration $\dot{\phi} - \dot{f}$ when the deviation of the rotation rate from orbital motion is smaller than the orbital motion, and to assume that the lag angle is constant when the rotation is much faster than orbital motion. An alternative of constant time-lag tides (for all frequencies) is possible but leads to very large lag angles and very fast tidal accelerations when early Earth had a fast spin⁴⁷; our even higher AM initial conditions would make this problem worse.

In order to treat tides on the Moon and Earth uniformly, including the obliquity-related terms, we implemented the following relationship between the tidal Q (inverse of the lag angle) and frequency

$$Q = Q_0 \sqrt{1 + \frac{1}{\delta^2} - \frac{1}{\delta}} \quad (20)$$

with the parameter δ defined as:

$$\delta = |\mathbf{w} - \mathbf{v}_t|/|\mathbf{v}_t| \quad (21)$$

where \mathbf{v}_t is the tangential component of the perturber's apparent velocity relative to the perturbee's (that is, the deformed body's) centre of mass, and the vector \mathbf{w} is defined as $\mathbf{w} = \mathbf{s} \times \mathbf{r}$, with \mathbf{s} being the angular velocity vector of the perturbee, and \mathbf{r} the radius-vector of the perturber relative to the perturbee. This gives us $Q = Q_0/\delta$ for small librations, and $Q = Q_0$ for a fast-rotating body, and a reasonable transition when $\delta \approx 1$. The way this lag angle is implemented in the integrator is to place a prolate quadrupole moment on the perturbee, with the axis of symmetry defined by the vector $\hat{\mathbf{r}} + \hat{\delta}(2Q)^{-1}$ (where hats denote unit vectors). This quadrupole moment (described by equation 4.145 of ref. 39) then torques (and is torqued by) the perturber, producing tidal accelerations. On the Moon we have only the tidal bulge raised by Earth but on Earth we have both lunar and solar tidal bulges, each of which torques both the Sun and the Moon; Moon–Sun tidal cross-terms have been found to be important^{12,48}.

Since we treated all orientations of relative motions between the orbital motion and the perturbee's velocity equally, this formulation does not discriminate between (eccentric) libration tides and obliquity tides. Our tests show that the integrator correctly reproduces the expected relationship¹⁹

$$\frac{di}{dt} = \frac{1 \sin^2 \theta}{7 \epsilon \tan i} \frac{d\theta}{dt} \quad (22)$$

Our 'levelling off' of the tidal phase lag for libration rates equal to about half of the orbital motion has little bearing on eccentricity evolution, given that lunar eccentricities in our scenario are rarely above 0.2. However, we do encounter large obliquities around the Cassini state transition, which means that we may be underestimating inclination damping at high obliquities when compared to analytical calculations^{19,31}. We think that our approach is justified, because stronger dissipation at large obliquities would require tidal lag angles much larger than that corresponding to the tidal Q in the fast-rotation case, which we treat as the upper limit on the lag angle. Further work may be needed to reconcile different definitions of tidal Q s in different treatments of high-obliquity tides.

Laplace plane transition. In Fig. 1, where a fast-spinning Earth's obliquity was set to 70°, we observe the instability associated with the Laplace plane transition when the Moon is at 17–18 R_E . During this transition the Moon acquires large inclination, which is similar to the findings of ref. 26. We find that the behaviour of lunar eccentricity is rather complex, exhibiting spikes and crashes, separated by relatively stable periods of moderately excited eccentricity. To understand these quasi-stable states, we looked in detail at the results of the simulation at 12.8 Myr (Extended Data Fig. 3). It is clear that the eccentricity is excited by Kozai-type perturbations^{24,49}, which are secular and involve interaction between an outer perturber (the Sun in this case) and an inclined perturbee (the Moon). Unlike in standard Kozai dynamics of planetary satellites^{50,51}, Earth's obliquity has an important role, and the inclination oscillates owing to mutual precession of the lunar orbit and Earth's spin axis, so that the inclination and eccentricity are not simply anti-correlated. Extended Data Figure 3 shows that this dynamical state is periodic in its secular behaviour, with exactly three inclination cycles for every eccentricity cycle. We hypothesize that each of the stable intervals in eccentricity (Figs 1 and 3, and Extended Data Figs 5 and 6) represent a periodic secular state where certain precession periods are locked in resonance.

Extended Data Figure 4 shows a different 'slice' of the simulation from Fig. 1 for the $Q_E/k_{2,E} = 200$ branch of the simulation (black line) at 34.6 Myr. At this time, inclination is about 30° and the Laplace plane is dominated by the Sun. The eccentricity is still excited, but not by Kozai-type perturbations. The source of eccentricity excitation is near-resonant perturbations stemming from the slow-varying argument $\Psi = 3\Omega + 2\omega - 3\gamma$, where Ω and γ are the longitudes of the

lunar ascending node and Earth's vernal equinox, respectively, and ω is the Moon's argument of perigee. This interaction term combines solar perturbations and the precession of Earth's spin axis, so it is capable of changing the obliquity of Earth. Since this interaction does not require capture into resonance, it is probably less sensitive to random perturbations such as planetesimal encounters³ than a narrow resonance would be. We hope to study the dynamics of this and similar near-resonant terms in future work.

Extended Data 5 shows the evolution of the Earth–Moon system with the same initial conditions as in Fig. 1, but with $Q_E/k_{2,E} = 200$ from the beginning of the simulation. The black line plots the case with $Q_M/k_{2,M} = 200$, and the grey line represents $Q_M/k_{2,M} = 50$. We also ran a simulation with $Q_M/k_{2,M} = 100$, and it was in all ways intermediate between these two. Overall, the simulations shown in Extended Data Fig. 5 have low final obliquities, but have an excess of AM compared with the real Earth–Moon system. We conclude that higher Q_E during the early part of the Laplace plane transition leads to less AM loss than in the case shown in Fig. 1, but that it leads to larger reduction in obliquity during the later part of the simulation. Lunar tidal properties appear to be less important, with a more dissipative moon leading to slightly more AM loss and lower final obliquities for Earth.

In Extended Data Fig. 6 we started Earth with a spin period of only 2 h (as opposed to 2.5 h in Fig. 1). Just like in Fig. 1, we changed the tidal properties of Earth halfway through the simulation, leading to two different outcomes (see legend to Extended Data Fig. 6). While the qualitative evolution of the system is similar to that shown in Fig. 1, there are some qualitative differences. The final obliquity is lower than in Fig. 1 (in excellent agreement with ref. 27), while the final AM is about 15% too large. Given that the trend in outcome we see with increasing initial AM (lower final obliquity and higher final AM) is in the opposite direction from the trend we see with increasing initial obliquity (higher final obliquity and lower final AM), it is likely that a more complete exploration of initial conditions will find higher-AM cases where both the final AM and obliquity are satisfactory.

The apparent threshold obliquity for the instability that we find (between 60° and 65°) is below the critical obliquity of about 69° found by ref. 23; we understand that this is because of non-zero initial inclination of the Moon when encountering the Laplace plane transition. In all of our simulations, the Moon acquires a few degrees of inclination when crossing the inclination resonance just interior to the evection⁶; this early resonance is weak at low obliquities but strong when Earth's spin axis is tilted. As long as we assume a hydrostatic shape for the Moon, the Moon's obliquity is low and obliquity tides are weak, so this inclination largely survives until the Laplace plane transition.

During the Laplace plane transition a large amount of energy is dissipated in the Moon. For example, during the evolution shown in Fig. 1, tidal heating in the Moon reaches 10¹⁴–10¹⁵ W for several tens of millions of years. Owing to the prolonged nature of both the transition and subsequent cooling of the lunar mantle, the lunar crust would contain the signal of major thermal events that occurred tens of millions of years after lunar accretion; thus, the Laplace plane transition should be considered when interpreting the geochronology of lunar samples.

Cassini state transition. The rotation states of most large planetary satellites are highly evolved through tidal forces raised by their parent planet. The spin periods of satellites are typically synchronized with the orbital period, and this is true for Earth's Moon. Apart from spin period synchronization, tidal forces also damp any initial, so-called 'free', obliquities of the moons. The most energetically stable final state for satellite obliquities is one where the spin axis maintains a constant angle to the moon's orbital plane. Since both the satellite's spin axis and the orbital plane are precessing (the spin axis precesses around the orbital plane, which precesses around the Laplace plane), the most stable final obliquity is one matched to inclination and the precession rates in such a way that the spin axis, orbit normal and Laplace plane normal all stay in the same plane during their precession. This arrangement is called a Cassini state, and Cassini states can be calculated using equation (7)³³. The two solutions relevant here are Cassini state 1, which the Moon will occupy when its spin precession rate is faster than the orbital precession, and Cassini state 2, which is the only possible solution when the spin precession is slower than orbital precession (as is the case at the present day). The Moon is thought to have crossed the Cassini state transition at about 33 R_E when Cassini state 1 ceased to exist and the Moon had to shift to Cassini state 2 (ref. 18).

The largest discrepancy between our R-SISTEM integration of the Cassini state transition (Fig. 4) and the semi-analytical solutions (Extended Data Fig. 1) is the fact that the Moon is in non-synchronous rotation from 29.7 R_E to about 35 R_E in the numerical simulation. To explore this phenomenon in more detail, we ran 516 short simulations with initial conditions on a grid in a and i covering the Cassini state transition (Extended Data Fig. 7). We found that the Moon settles into a stable sub-synchronous rotation for a wide range of initial conditions with $a = 27$ –37 R_E . In all of the grid simulations, the Moon was initially non-synchronous, so our results show when the synchronous rotation will not be re-established after being broken, rather than that the synchronous rotation is unstable, in the

regions covered with orange crosses in Extended Data Fig. 7. When it comes to intrinsic stability of continued synchronous rotation, we find major differences between Cassini states 1 and 2. Cassini state 1 is by itself stable all the way until its disappearance at $a = 31\text{--}34R_E$ (the exact distance depends on inclination), unless disturbed by an outside influence (for example, a wobble resonance or an impact). On the other hand, we find that Cassini state 2 is intrinsically unstable if its equilibrium obliquity is above 58.15° (Extended Data Fig. 8), a result well established in the literature^{52,53}, but not previously relevant for the Moon in the low-inclination case. We also find that once Cassini state 2 becomes stable, the Moon may still occupy a sub-synchronous rotation state just short of synchronous until it is disturbed by a resonance or an impact.

We also find a number of resonances that excite the Moon's rotation and can lead to long-term chaos (red crosses in Extended Data Fig. 7). Some of these are probably associated with secondary resonances found by ref. 54. Here we will concentrate on three features seen in Fig. 4: the wobble resonance at 29.7° , the 1:2 spin-orbit resonance at 31° and several resonances at $34\text{--}35^\circ$. The wobble resonance at 29.7° breaks the synchronous rotation by inducing large non-principal axis rotation within the Moon. Extended Data Fig. 9 shows how the wobble amplitude grows as the Moon approaches the resonance (at a tidal evolution rate accelerated 100 times), and Extended Data Fig. 10 shows the resonance crossing itself, integrated at the nominal tidal evolution rate. This resonance is caused by the commensurability between Earth's heliocentric motion and the lunar libration in longitude, the frequency of which is given by ref. 55

$$\lambda = \sqrt{\frac{3(B-A)}{C}} n \quad (23)$$

Currently, the period of the libration in longitude is 1,056.1 days (ref. 55). As long as the Moon's shape is constant, this period is proportional to the orbital period, so the semimajor axis at which this period is one year is

$$a_r = \left(\frac{365.25 \text{ days}}{1056.1 \text{ days}} \right)^{2/3} 60.3R_E = 29.7R_E \quad (24)$$

which is exactly where it occurs in our numerical simulation. After the wobble disrupts synchronous rotation, the Moon settles into a sub-synchronous high-obliquity state, consistent with Extended Data Figs 7 and 8.

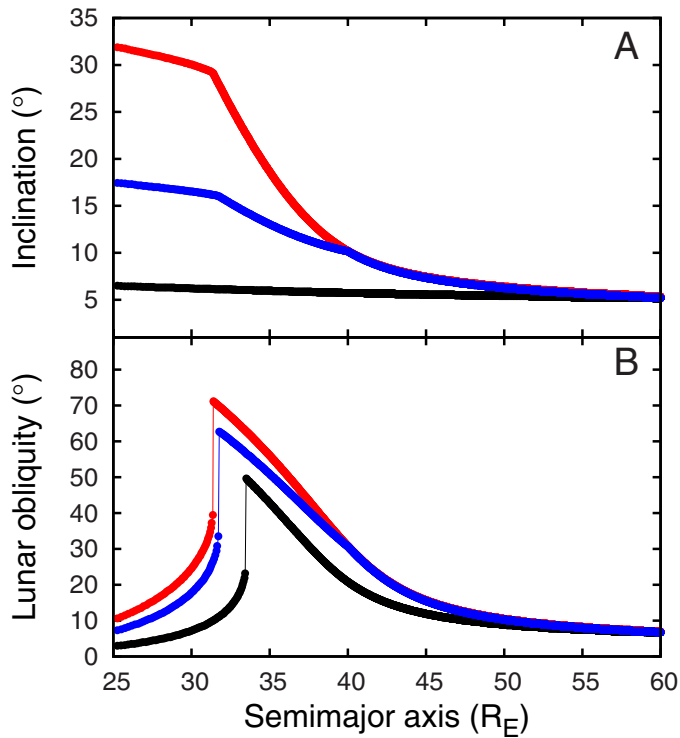
At about $31R_E$, the Moon is briefly captured into a 1:2 spin-orbit resonance. This resonance can be stable at high eccentricities⁵⁶, but we never observed long periods of capture in our high-obliquity, low-eccentricity cases. At $34\text{--}35R_E$ the Moon encounters several resonances which disrupt sub-synchronous rotation and the Moon settles into (the by now stable) Cassini state 2. We think that these resonances are related to the 1:3 secondary secular resonance, which was found to intersect with Cassini state 2 at about this geocentric distance⁵⁴. Clearly more work is needed to identify these resonant features and explore the full diversity of the Moon's past spin-orbit dynamics. Here we assumed that the Moon already had its present shape at the time of Cassini state transition, which is a reasonable assumption³⁰, but other shape histories are possible. Also, beyond $30R_E$ in Fig. 4, lunar tidal evolution is accelerated 10 times over nominal, and the Moon would probably be affected by some of the later resonances more strongly if the evolution was integrated at the nominal rate. Additional factors that we ignored are impacts that can disrupt the Moon's rotational state⁵⁶ and the core-mantle interaction, which may have been important in generating the ancient lunar magnetic field⁵⁷.

Once the Cassini state is re-established in Fig. 4, we can compare our numerical results with analytical estimates (see equations (1)–(6)). Our numerical code R-SISTEM damps lunar inclination from 16° at $35R_E$ to 9.2° at $40R_E$. This is slightly below our nominal target of 10° at $40R_E$ (based on Extended Data Fig. 1), obtained by assuming a long-term-average Earth and the present Moon beyond $40R_E$ (Extended Data Fig. 1). Also, R-SISTEM damps inclination slightly faster

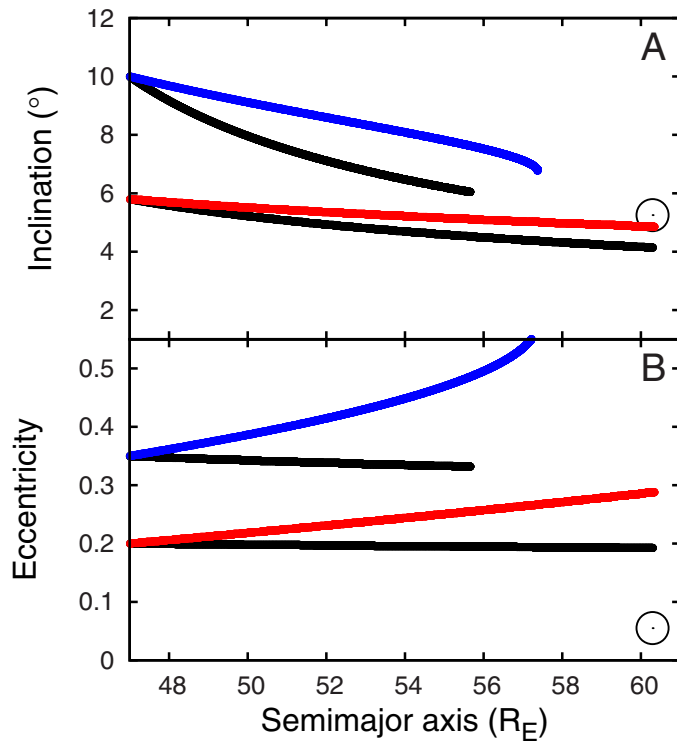
for $Q_{M,0} = 60$ than does the analytical model (switching to the analytical approach at $35R_E$ would give us 10.9° at $40R_E$). This is an inevitable consequence of the differences between the two approaches, but we are encouraged by the overall convergence of the results. In the absence of any strong constraints on the timeline of the Moon's tidal evolution, we think that the history shown in Figs 1 and 4 represents the best currently available explanation for the otherwise puzzling lunar orbital inclination.

Code Availability. Computer programs used in this paper are available on request.

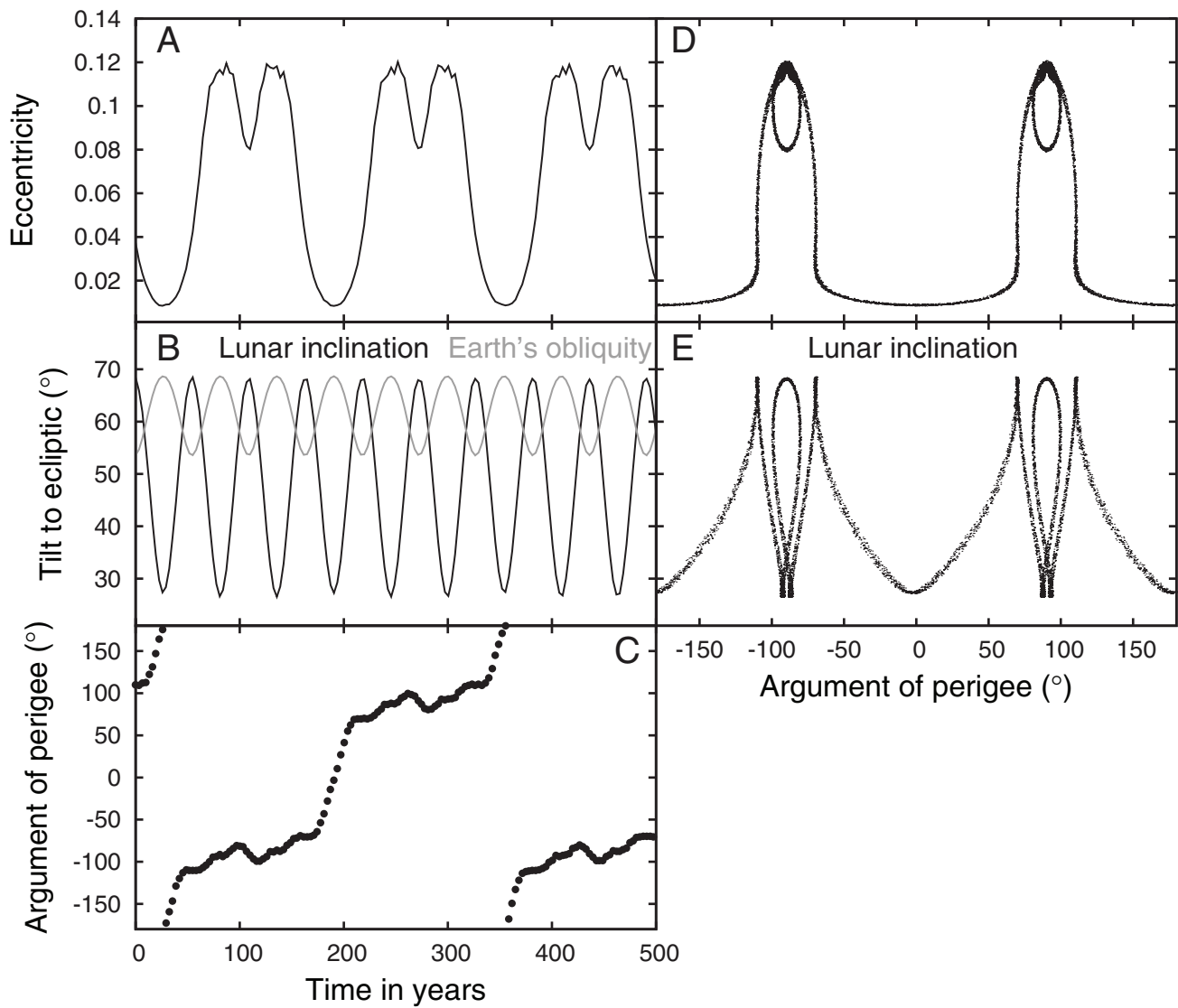
31. Peale, S. J. & Cassen, P. Contribution of tidal dissipation to lunar thermal history. *Icarus* **36**, 245–269 (1978).
32. Canup, R. M., Levison, H. F. & Stewart, G. R. Evolution of a terrestrial multiple-moon system. *Astron. J.* **117**, 603–620 (1999).
33. Peale, S. J. Generalized Cassini's Laws. *Astron. J.* **74**, 483 (1969).
34. Ćuk, M. & Burns, J. A. On the secular behavior of irregular satellites. *Astron. J.* **128**, 2518–2541 (2004).
35. Williams, J. G. & Boggs, D. H. Tides on the Moon: theory and determination of dissipation. *J. Geophys. Res. Planets* **120**, 689–724 (2015).
36. Webb, D. J. Tides and the evolution of the Earth–Moon system. *Geophys. J.* **70**, 261–271 (1982).
37. Bills, B. G. & Ray, R. D. Lunar orbital evolution: a synthesis of recent results. *Geophys. Res. Lett.* **26**, 3045–3048 (1999).
38. Ćuk, M. Excitation of lunar eccentricity by planetary resonances. *Science* **318**, 244 (2007).
39. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* (Cambridge Univ. Press, 1999).
40. Wisdom, J. & Holman, M. Symplectic maps for the N -body problem. *Astron. J.* **102**, 1528–1538 (1991).
41. Chambers, J. E., Quintana, E. V., Duncan, M. J. & Lissauer, J. J. Symplectic integrator algorithms for modeling planetary accretion in binary star systems. *Astron. J.* **123**, 2884–2894 (2002).
42. Ćuk, M., Dones, L. & Nesvorný, D. Dynamical evidence for a late formation of Saturn's Moons. *Astrophys. J.* **820**, 97 (2016).
43. Tourné, J. & Wisdom, J. Lie-Poisson integrators for rigid body dynamics in the solar system. *Astron. J.* **107**, 1189–1202 (1994).
44. Vokrouhlický, D., Breiter, S., Nesvorný, D. & Bottke, W. F. Generalized YORP evolution: onset of tumbling and new asymptotic states. *Icarus* **191**, 636–650 (2007).
45. Sharma, I., Burns, J. A. & Hui, C.-Y. Nutational damping times in solids of revolution. *Mon. Not. R. Astron. Soc.* **359**, 79–92 (2005).
46. Burns, J. A. Elementary derivation of the perturbation equations of celestial mechanics. *Am. J. Phys.* **44**, 944–949 (1976).
47. Meyer, J., Elkins-Tanton, L. & Wisdom, J. Coupled thermal-orbital evolution of the early Moon. *Icarus* **208**, 1–10 (2010); corrigendum *Icarus* **212**, 448–449 (2011).
48. Laskar, J. et al. A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
49. Lidov, M. L. The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planet. Space Sci.* **9**, 719–759 (1962).
50. Innanen, K. A., Zheng, J. Q., Mikkola, S. & Valtonen, M. J. The Kozai mechanism and the stability of planetary orbits in binary star systems. *Astron. J.* **113**, 1915 (1997).
51. Carruba, V., Burns, J. A., Nicholson, P. D. & Gladman, B. J. On the inclination distribution of the Jovian irregular satellites. *Icarus* **158**, 434–449 (2002).
52. Beletskii, V. V. Resonance rotation of celestial bodies and Cassini's laws. *Celestial Mech.* **6**, 356–378 (1972).
53. Gladman, B., Quinn, D. D., Nicholson, P. & Rand, R. Synchronous locking of tidally evolving satellites. *Icarus* **122**, 166–192 (1996).
54. Wisdom, J. Dynamics of the lunar spin axis. *Astron. J.* **131**, 1864–1871 (2006).
55. Rambaux, N. & Williams, J. G. The Moon's physical librations and determination of their free modes. *Celestial Mech. Dyn. Astron.* **109**, 85–100 (2011).
56. Wiczkorek, M. A., Correia, A. C. M., Le Feuvre, M., Laskar, J. & Rambaux, N. Mercury's spin-orbit resonance explained by initial retrograde and subsequent synchronous rotation. *Nat. Geosci.* **5**, 18–21 (2011).
57. Meyer, J. & Wisdom, J. Precession of the lunar core. *Icarus* **211**, 921–924 (2011).



Extended Data Figure 1 | Semi-analytical model of the lunar tidal evolution. Evolution of lunar inclination (a) and obliquity (b) as the Moon evolves from $25R_E$ to $60R_E$ using our semi-analytical model (Methods subsection ‘Damping of lunar inclination by obliquity tides’). Initial inclinations were chosen so that the final lunar inclination was the current value of about 5° , while the obliquity was calculated assuming the Moon was in a Cassini state (jumps between $30R_E$ and $35R_E$ are due to the transitions between Cassini states 1 and 2)¹⁸. Love numbers were set at their current values ($k_{2,E} = 0.3$, $k_{2,M} = 0.024$)³⁵, and the current lunar shape was assumed. The black and red lines plot the solutions for $Q_M = 10,000$ and $Q_M = 38$ (current value), respectively, while Q_E was in the range 33–35 range (it was adjusted so that a semimajor axis of $60R_E$ was reached after 4,500 Myr). The blue line plots a history assuming $Q_M = 100$ interior to $40R_E$, and $Q_M = 38$ after the Moon passes that distance. The black line closely resembles the results of studies^{11,12} that neglected lunar obliquity tides, while the other two curves indicate that the past lunar inclination must have been much larger owing to lunar obliquity tides.

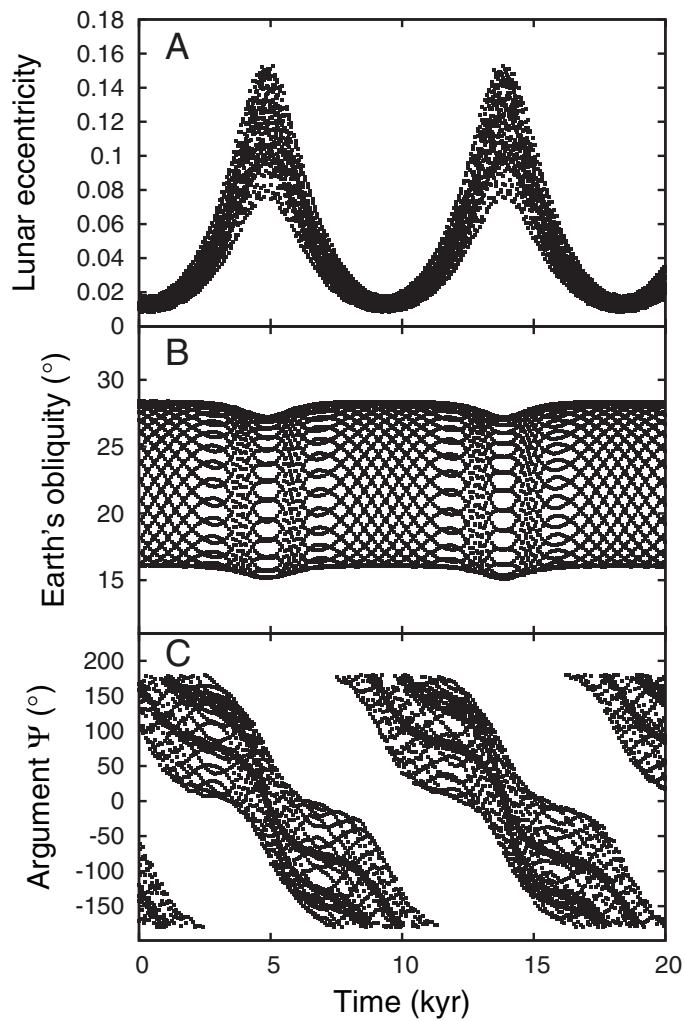


Extended Data Figure 2 | Lunar tidal evolution following planetesimal encounters. Evolution of lunar inclination (a) and eccentricity (b) following the excitation of the lunar inclination by encounters with planetesimals as proposed by ref. 3, using our semi-analytical model. The two sets of initial conditions are for the state $a = 47R_E$, $i = 5.8^{\circ}$ featured in ref. 3 (Fig. 1), and a possible outcome with a more excited inclination $i = 10^{\circ}$ (also at $a = 47R_E$). The initial eccentricities were estimated as $e = 2\sin i$. The black lines show the evolution assuming $Q_M = 38$ and $Q_E = 34$, with the current Love numbers. The red line plots the evolution for $Q_M = 100$ and the blue line the evolution for $Q_E = 20$. The circle symbol plots the current inclination and eccentricity of the lunar orbit. No combination of tidal parameters can simultaneously match both the current lunar inclination and eccentricity at the same time. A small Q_M combined with high e also keeps the Moon from reaching $60R_E$ (top two lines). Decreasing Q_E also does not help, as stronger Earth tides further increase the lunar eccentricity (blue line).

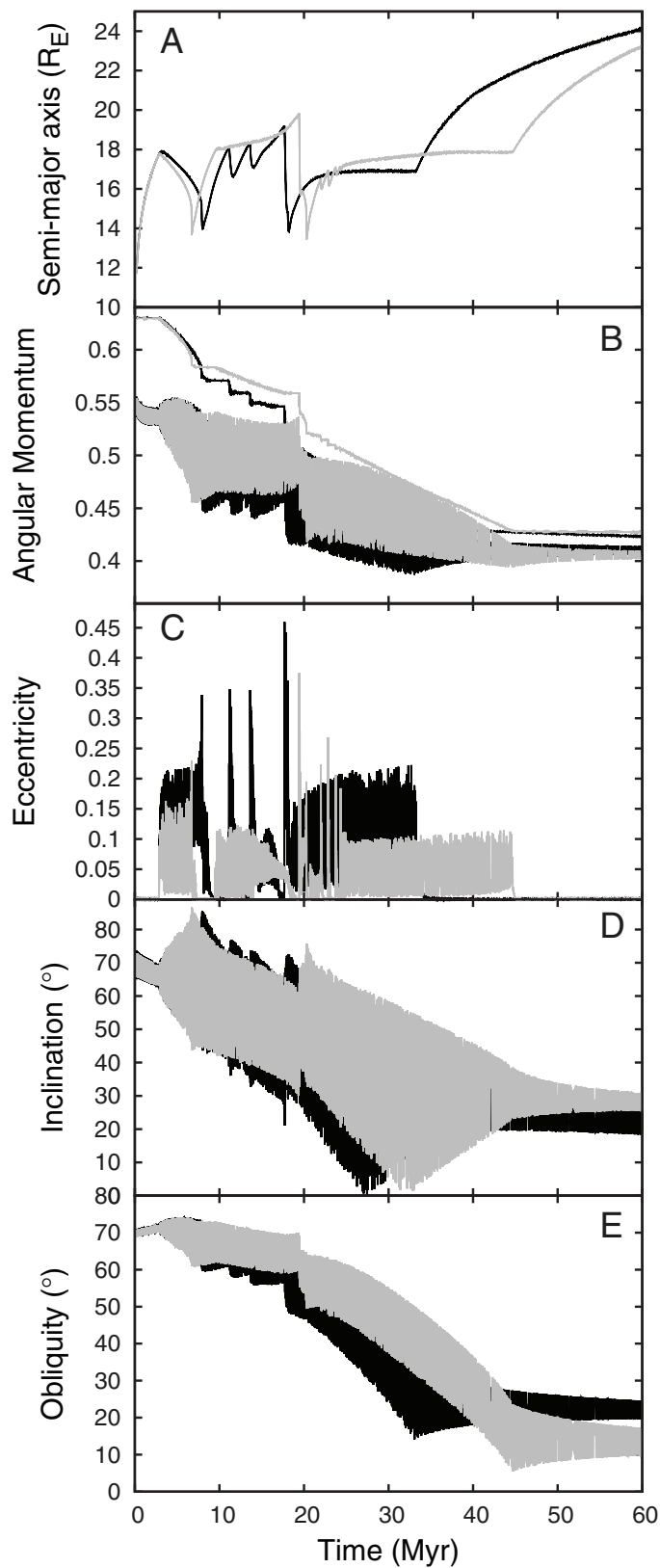


Extended Data Figure 3 | A snapshot of the simulation shown in Fig. 1 taken at 12.8 Myr. The left-hand panels show eccentricity (a), lunar inclination and Earth's obliquity with respect to the ecliptic (b), and the argument of perigee of lunar orbit, with the ecliptic as fundamental plane (c) versus time over a 500-year period, while the right-hand panels plot lunar eccentricity (d) and inclination (e) against the Moon's argument of perigee (over the whole period of 30,000 yr). The eccentricity is clearly

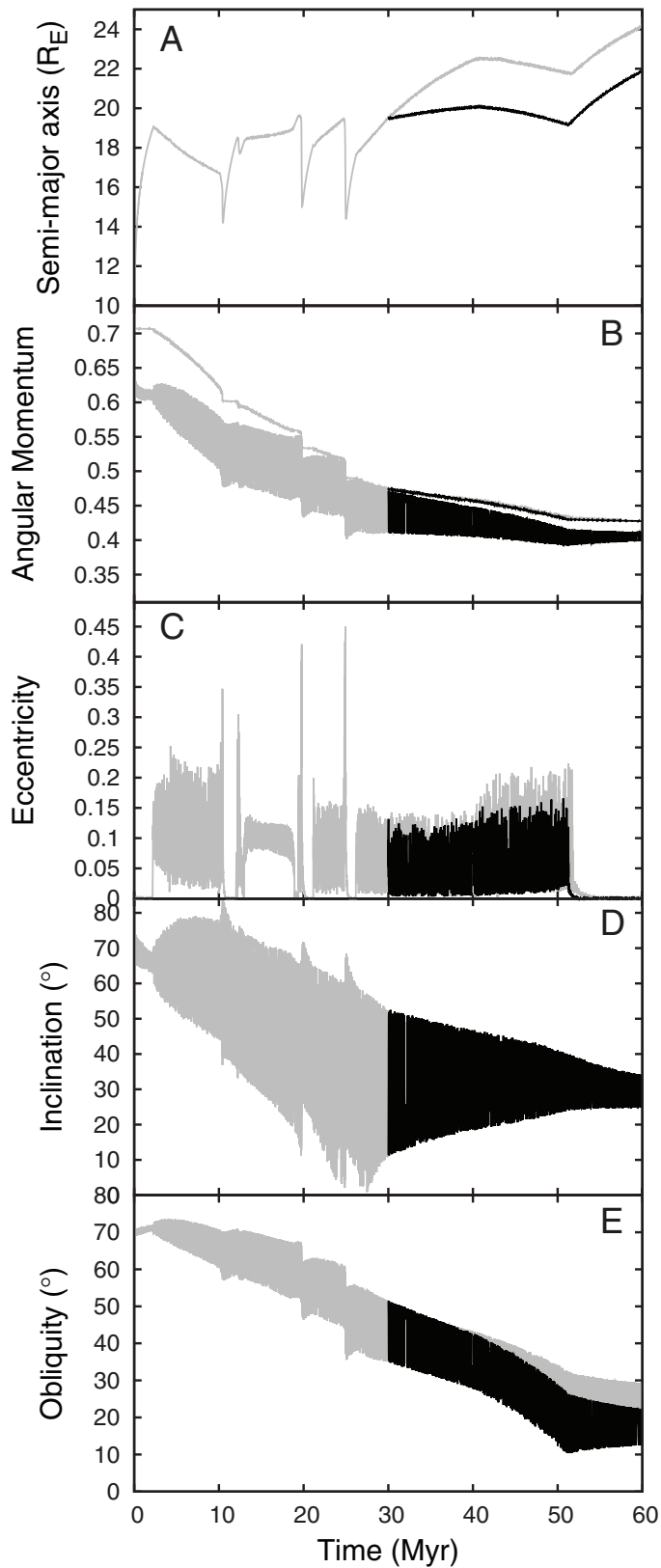
correlated with the argument of perigee, as expected for Kozai-type perturbations^{24,49}. Rapid mutual precession of Earth's spin axis and the plane of the Moon's orbit, which are substantially inclined to one another (and to the ecliptic plane), clearly affects both the eccentricity and the perigee precession. This secular behaviour is periodic, with exactly three inclination cycles per one eccentricity cycle, which corresponds to half of the period of precession of the argument of perigee.



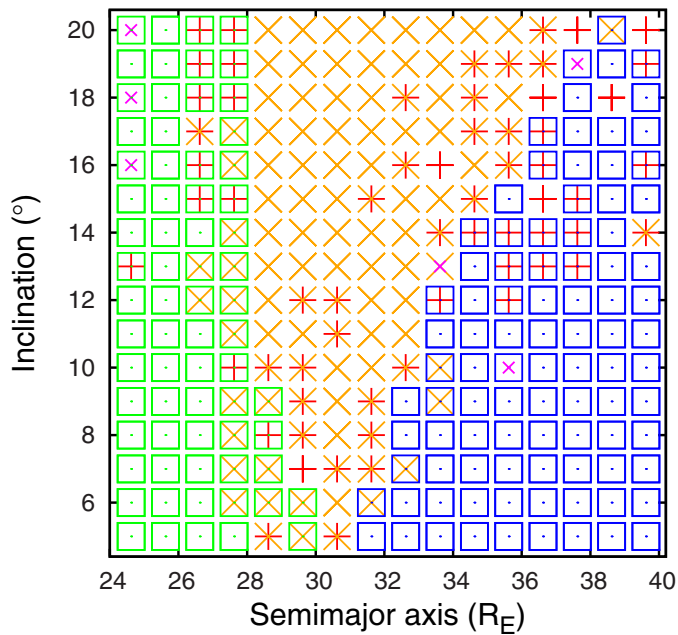
Extended Data Figure 4 | A snapshot of $Q_E/k_{2,E} = 200$ simulation shown in Fig. 1 (black line) at 34.6 Myr. At this time, eccentricity excitation (a) and the associated variation of Earth's obliquity (b) are not due to Kozai perturbations, but owing to the slow-varying near-resonant argument $\Psi = 3\Omega + 2\omega - 3\gamma$ (c), where Ω and γ are longitudes of the lunar ascending node and Earth's vernal equinox, respectively, and ω is the Moon's argument of perigee. This near-resonant interaction is responsible for the substantial reduction of Earth's obliquity seen in Fig. 1.



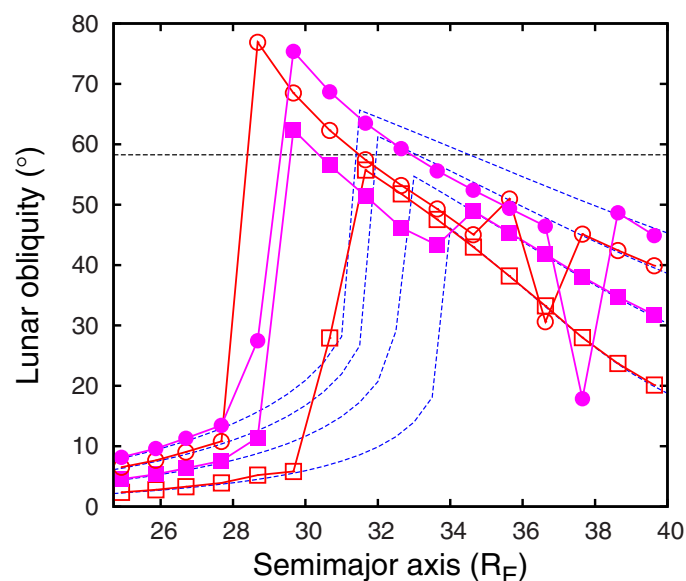
Extended Data Figure 5 | Early tidal evolution of the Moon with $Q_E/k_{2,E} = 200$ throughout the simulations. Black lines are for the case with $Q_M/k_{2,M} = 200$, while grey lines plot the simulation with $Q_E/k_{2,E} = 50$. The most notable aspects of these simulations are low final obliquities of Earth (e) and a final AM of the Earth–Moon system (b) in excess of the current value of $0.35\alpha_E\sqrt{GM_E^3R_E}$. a, c and d plot the semimajor axis, eccentricity and inclination of the Moon.



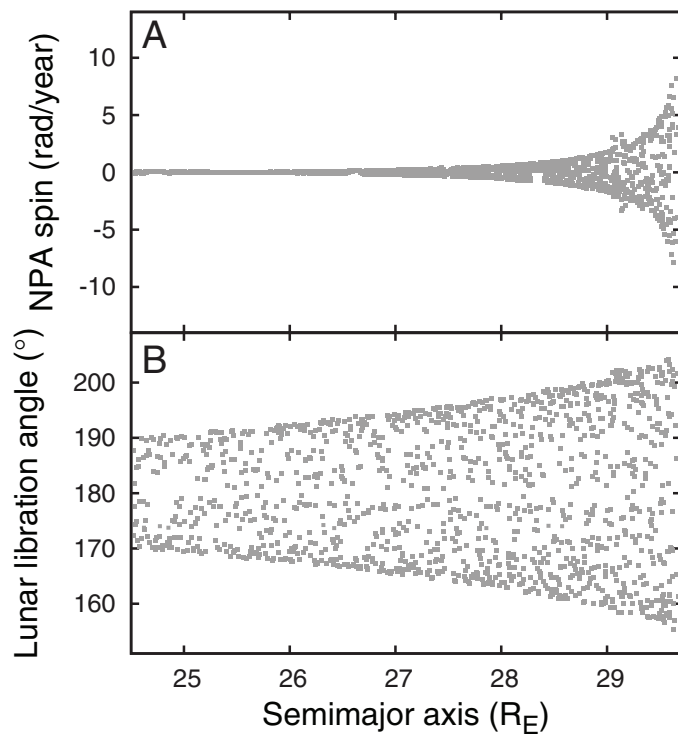
Extended Data Figure 6 | Early tidal evolution of the Moon with Earth initially having a 2-h spin period. This is equivalent to the system having twice the current AM. The grey lines plot a simulation in which the tidal properties of Earth and the Moon were $Q_E/k_{2,E} = Q_M/k_{2,M} = 100$ throughout. The black line shows a simulation branching at 30 Myr by changing $Q_E/k_{2,E}$ to 200. Although the final obliquity of Earth (e) is correct, the final AM of the Earth–Moon system (b) somewhat exceeds the current value of $0.35 \times \alpha_E \sqrt{GM_E^3 R_E}$. a, c and d plot the semimajor axis, eccentricity and inclination of the Moon.



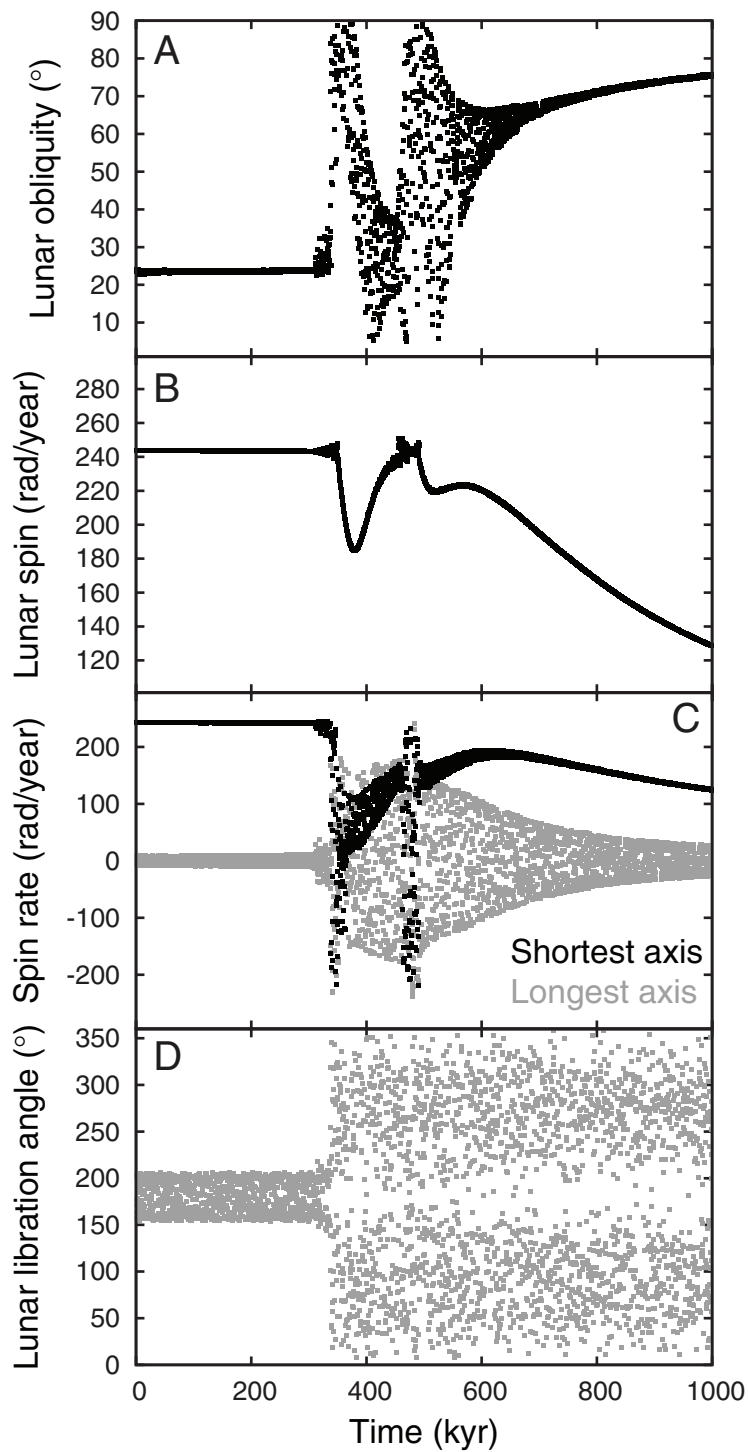
Extended Data Figure 7 | Map of lunar rotational dynamics close to the Cassini state transition. Outcomes of 512 simulations probing the end states of initially very fast and very slow lunar rotations for 16 different lunar semimajor axes a and 16 different lunar inclinations i . Simulations were run for 1 Myr, except for the rightmost three columns, which were followed for 3 Myr. Each $a-i$ field is described by two symbols, one each for initial rotations of 127 rad yr^{-1} and 381 rad yr^{-1} . Green and blue boxes indicate synchronous rotation in Cassini states 1 and 2, respectively. Crosses indicate non-synchronous rotation with stable obliquity, with large orange crosses indicating sub-synchronous rotation, and small magenta crosses plotting super-synchronous states. Red crosses signify variations in obliquity above 1° during the last 50 kyr of the simulation (indicating excited or chaotic spin axis precession).



Extended Data Figure 8 | Lunar obliquity close to the Cassini state transition. Obliquities for four 'slices' in inclination (at 5°, 10°, 15° and 20°) from the grid of short simulations shown in Extended Data Fig. 7 (solid red and magenta lines with points; the obliquities and inclinations are in the same order at far left and far right). When two different simulations for the same a and i differed in outcome, we chose the solution within the Cassini state, if available. The blue dashed lines plot the relevant Cassini states calculated using analytical formulae, while the black dashed line at 58.15° plots the upper limit for stable obliquities in the relevant Cassini state. Although the numerical and analytical results agree at the smallest and largest semimajor axes, the large discrepancies in between are due to non-synchronous rotations being dominant at the Cassini state transition.



Extended Data Figure 9 | The Moon's wobble as it approaches the annual resonance in Fig. 4. The rotation rate around the longest axis of the Moon (a) and the angle between the longest axis and Earth (b) during the first phase of lunar tidal evolution (red points in Fig. 4) within $29.7R_E$, where we accelerated the tidal evolution by a factor of a hundred. The wobble is clearly building up as the Moon is approaching the resonance between its free wobble and Earth's orbital period at about $29.7R_E$. The growth in lunar libration angle is more influenced by increasing lunar obliquity (Fig. 4b) than the increase in the amplitude of the wobble.



Extended Data Figure 10 | Passage through the annual resonance of the lunar free wobble in Fig. 4. Lunar obliquity (a), spin rate (b), rotation rates around the longest and shortest principal axes (c), and the angle between the Moon's longest axis and Earth (d) during the first 1 Myr of the 'blue' segment of lunar tidal evolution in Fig. 4 (which was simulated at the nominal rate for tidal evolution). The free wobble (tracked by grey points in c) experiences a resonance at about 330 kyr, breaking the Moon's synchronous rotation. Since the Moon is close to the Cassini state transition, it cannot evolve back into Cassini state 1 and it settles into a non-synchronous high-obliquity state⁵⁸.

Nanoscale thermal imaging of dissipation in quantum systems

D. Halbertal¹, J. Cuppens^{1,2}, M. Ben Shalom^{3,4}, L. Embon^{1,†}, N. Shadmi⁵, Y. Anahory¹, H. R. Naren¹, J. Sarkar¹, A. Uri¹, Y. Ronen¹, Y. Myasoedov¹, L. S. Levitov⁶, E. Joselevich⁵, A. K. Geim^{3,4} & E. Zeldov¹

Energy dissipation is a fundamental process governing the dynamics of physical, chemical and biological systems. It is also one of the main characteristics that distinguish quantum from classical phenomena. In particular, in condensed matter physics, scattering mechanisms, loss of quantum information or breakdown of topological protection are deeply rooted in the intricate details of how and where the dissipation occurs. Yet the microscopic behaviour of a system is usually not formulated in terms of dissipation because energy dissipation is not a readily measurable quantity on the micrometre scale. Although nanoscale thermometry has gained much recent interest^{1–15}, existing thermal imaging methods are not sensitive enough for the study of quantum systems and are also unsuitable for the low-temperature operation that is required. Here we report a nano-thermometer based on a superconducting quantum interference device with a diameter of less than 50 nanometres that resides at the apex of a sharp pipette: it provides scanning cryogenic thermal sensing that is four orders of magnitude more sensitive than previous devices—below $1\ \mu\text{K Hz}^{-1/2}$. This non-contact, non-invasive thermometry allows thermal imaging of very low intensity, nanoscale energy dissipation down to the fundamental Landauer limit^{16–18} of 40 femtowatts for continuous readout of a single qubit at one gigahertz at 4.2 kelvin. These advances enable the observation of changes in dissipation due to single-electron charging of individual quantum dots in carbon nanotubes. They also reveal a dissipation mechanism attributable to resonant localized states in graphene encapsulated within hexagonal boron nitride, opening the door to direct thermal imaging of nanoscale dissipation processes in quantum matter.

Investigation of energy dissipation on the nanoscale is of major fundamental interest for a wide range of disciplines, ranging from biological processes, through chemical reactions, to energy-efficient computing^{1–5}. Study of dissipation mechanisms in quantum systems is of particular importance because dissipation demolishes quantum information. In order to preserve a quantum state, the dissipation has to be extremely weak and hence hard to measure. As a figure of merit for detection of low-power dissipation in quantum systems¹⁶, we consider an ideal qubit operating at a typical read-out frequency of 1 GHz. Landauer's principle states that the lowest bound on energy dissipation in an irreversible qubit operation is $E_0 = k_B T \ln 2$, where k_B is Boltzmann's constant and T is the temperature^{17,18}. At $T = 4.2\ \text{K}$, $E_0 = 4 \times 10^{-23}\ \text{J}$, several orders of magnitude below both the $10^{-19}\ \text{J}$ of dissipation per logical operation in present-day superconducting electronics and the $10^{-15}\ \text{J}$ in CMOS devices^{19,20}. Hence the power dissipated by an ideal qubit operating at a read-out rate of $f = 1\ \text{GHz}$ will be as low as $P = E_0 f = 40.2\ \text{fW}$. The resulting temperature increase of the qubit will depend on its size and the thermal properties of the substrate. For example, a $120\ \text{nm} \times 120\ \text{nm}$ device on a $1\text{-}\mu\text{m}$ -thick SiO_2/Si substrate dissipating $40\ \text{fW}$ will heat up by about $3\ \mu\text{K}$ (Fig. 1). Such signals are several orders of magnitude below the best sensitivity

(several $\text{mK Hz}^{-1/2}$) of any of the existing imaging techniques^{1–15} (Fig. 1a): these include radiation-based thermometry using infrared⁶ (IR), fluorescence in nanodiamonds^{4,7,8}, Raman spectroscopy⁹, scanning near-field optical microscopy¹⁰ (SNOM), or the electron beam induced plasmons in transmission electron microscopy⁵ (TEM), and atomic force microscopy (AFM) equipped with thermocouple (TC SThM) or resistive (Res. SThM) thermometers^{10–15}. Moreover, none of the existing imaging techniques has been demonstrated to operate at the low temperatures that are essential for study of quantum systems.

Superconducting junctions are commonly used as highly sensitive thermometers^{2,21}, relying on the strong temperature dependence of their critical current $I_c(T)$, which in the vicinity of the critical temperature T_c can be approximated by $I_c(T) \approx I_0(1 - T/T_c)$. Conventional junction configurations, however, are not suitable for scanning probe thermometry owing to their planar geometry and strong thermal coupling to the substrate.

Here we introduce a non-contact cryogenic scanning probe microscopy technique based on a superconducting quantum interference device (SQUID) positioned on a tip^{22,23}. We utilize a novel approach in which either a single Pb ($T_c = 7.2\ \text{K}$) superconducting junction or a SQUID is fabricated on the apex of a sharp quartz pipette (SQUID on tip, SOT), and can be made as small as a few tens of nanometres in diameter (Supplementary Information section S1). Figure 1b shows a scanning electron microscopy (SEM) image of a thermometer using this SQUID-on-tip technique (that is, a tSOT) with an effective diameter of $46\ \text{nm}$, as determined from its quantum interference pattern. Similar SOT devices have previously been reported^{22,23} to be extremely sensitive magnetic sensors, reaching a magnetic spin sensitivity below $0.4\ \mu_B\ \text{Hz}^{-1/2}$. The electrical characteristics of the tSOT (Fig. 1c) show the current through the sensor (I_{SOT}) versus externally applied bias current (I_{bias}) at various temperatures (see electrical diagram in Supplementary Fig. 1c). For $I_{\text{SOT}} < I_c$, essentially all the applied current flows through the tSOT, while at higher bias a substantial part of the current diverts to a parallel shunt resistor. When biasing the tSOT at $I_{\text{bias}} > I_c$ (dashed line in Fig. 1c), the temperature dependence of $I_{\text{SOT}}(T)$ gives rise to a thermal response $dI_{\text{SOT}}/dT = -9.5\ \mu\text{A K}^{-1}$ (Supplementary Fig. 1b); when this is combined with the very low white noise of the tSOT ($S_I^{1/2} = 8.3\ \text{pA Hz}^{-1/2}$; Supplementary Fig. 1c) it translates into a remarkably low thermal noise of $S_T^{1/2} = 870\ \text{nK Hz}^{-1/2}$, an improvement of four orders of magnitude over existing thermal imaging methods (Fig. 1a).

To enable effective thermal imaging, the thermal properties of the sensor, including its coupling to the sample, are crucial. For non-invasive imaging, the thermal resistance between the sensor and the investigated device, R_{sd} , has to be considerably larger than the thermal resistance between the device and the bulk of its substrate, R_{db} (Fig. 1d). On the other hand, in order for the temperature of the sensor T_s to accurately describe the local temperature of the device T_d , a high

¹Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 7610001, Israel. ²Catalan Institute of Nanoscience and Nanotechnology (ICN2), CSIC and the Barcelona Institute of Science and Technology, Campus UAB, Bellaterra, 08193 Barcelona, Spain. ³National Graphene Institute, The University of Manchester, Booth Street East, Manchester M13 9PL, UK. ⁴School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK. ⁵Department of Materials and Interfaces, Weizmann Institute of Science, Rehovot 7610001, Israel. ⁶Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present address: Department of Physics, Columbia University, New York, New York 10027, USA.

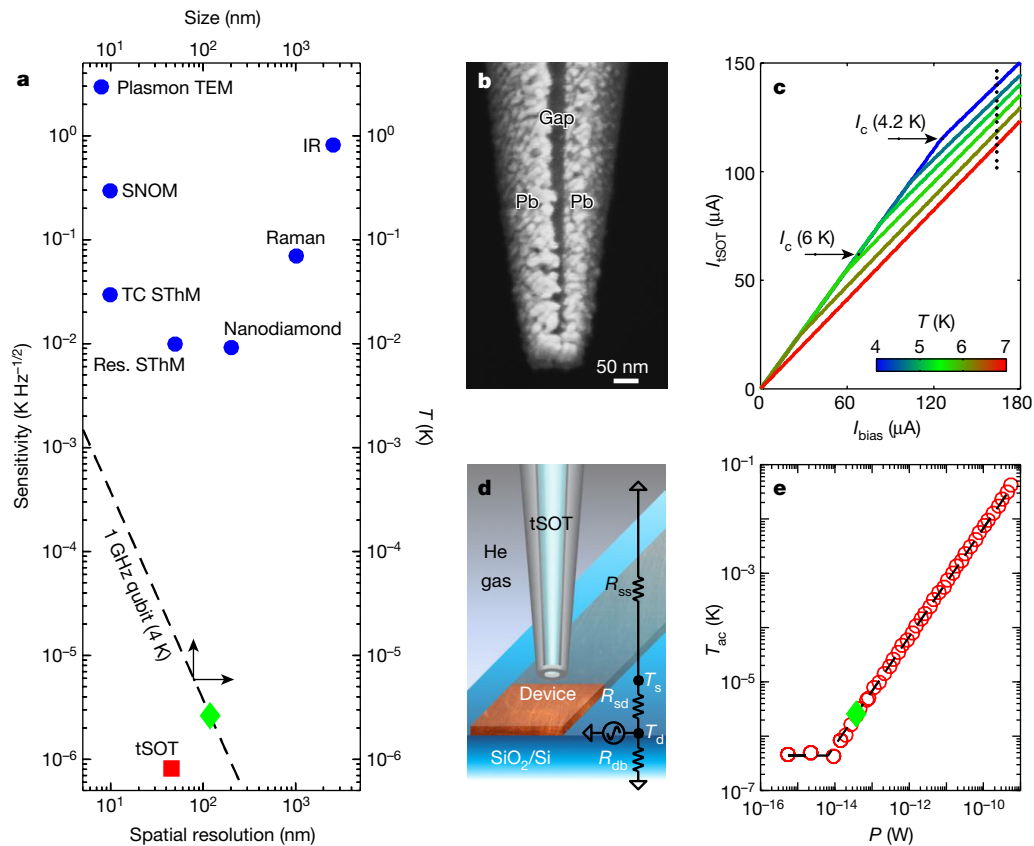


Figure 1 | tSOT characteristics and performance. **a**, Sensitivities (left-hand axis) of different thermal imaging techniques (blue) and of the tSOT (red) versus their spatial resolution (bottom axis). Green diamond, the measured temperature increase (right-hand axis) due to 40 fW dissipation (taken from **e**) corresponding to Landauer's limit for qubit operation at 4.2 K at 1 GHz along with the theoretical scaling of the temperature (right-hand axis) with the qubit size (dashed line, top axis). See main text for definitions of the thermal imaging techniques. **b**, SEM image of the 46-nm effective diameter Pb tSOT. **c**, Electrical characteristics of the tSOT

thermal resistance R_{ss} is required between the sensor and its support structure. The resulting overall requirement of $R_{ss} \gg R_{sd} \gg R_{db}$ is usually hard to achieve in AFM-type scanning thermal probes¹², leading to invasive in-contact imaging^{13,14}. In a tSOT, in contrast, R_{ss} is extremely high, owing to the unique nanoscale cross-section geometry of the device, giving rise to a quantum-limited phonon thermal conductivity²⁴, and to the absence of electronic heat conductivity along the superconducting leads. As a result, our smaller tSOT attains $R_{ss} \approx 10^{11} \text{ K W}^{-1}$ (see Supplementary Information section S6) as compared to $R_{db} \approx 10^7 \text{ K W}^{-1}$ for a $120 \text{ nm} \times 120 \text{ nm}$ device on a SiO_2/Si substrate at 4.2 K (Fig. 1e). The corresponding optimal R_{sd} of 10^8 to 10^{10} K W^{-1} is readily achieved in our configuration using a few mbar of He exchange gas to tune R_{sd} (Supplementary Fig. 8). These features permit non-contact sensing of the T_d of the sample with nanoscale resolution (Supplementary Fig. 2 and Supplementary Information sections S3, S4).

To characterize the thermal sensitivity of the scanning tSOT, we position the sensor above a 120-nm-wide Cu nanowire on a SiO_2/Si substrate (Fig. 1d and Supplementary Fig. 7); the nanowire is carrying an alternating current (a.c.), which results in an a.c. temperature modulation T_{ac} of the nanowire at 13.1 kHz. By changing the current amplitude, we measured tSOT T_{ac} versus the power P dissipated per square $120 \text{ nm} \times 120 \text{ nm}$ segment of the wire (Fig. 1e). Since $T_{ac} \ll T_b$, where $T_b = 4.2 \text{ K}$ is the thermal bath temperature, the measurement is in the small signal limit and hence T_{ac} is linear in P as expected, reaching a noise level of $T_{ac} \approx 440 \text{ nK}$ at $P \approx 6 \text{ fW}$. The green triangle in Fig. 1e shows the Landauer dissipation limit of $P = 40.2 \text{ fW}$ of a qubit operating

for temperatures ranging from 4.2 K to 7.2 K with marked values of the critical current I_c at representative temperatures. **d**, Schematic drawing of the measurement set-up and a simplified effective thermal circuit. See main text for definitions of symbols. **e**, Measurement of the tSOT temperature T_{ac} at 13.1 kHz above a 120-nm-wide Cu nanowire with sheet resistance 0.46Ω per square versus the a.c. power P dissipated by a variable I_{ac} in a $120 \text{ nm} \times 120 \text{ nm}$ unit segment of the nanowire. The dashed line shows a linear fit (with slope $R_{db} = 6.8 \times 10^7 \text{ K W}^{-1}$) and a noise floor of $T_{ac} \approx 440 \text{ nK}$ below $P \approx 6 \text{ fW}$ using a lock-in amplifier time constant of 1 s.

at 1 GHz at 4.2 K, which gives rise to $T_{ac} = 2.6 \mu\text{K}$ in our sample. This value is indicated in Fig. 1a (green triangle) along with the expected inverse scaling of the qubit temperature with its area (dashed line).

By applying a proper combination of magnetic field and bias, the tSOT can be tuned to have both magnetic field and thermal sensitivities. As the Oersted field generated by the transport current in the sample is linear in current while the dissipation is quadratic, the resulting magnetic field signal B_z^{ac} generated by the a.c. current I_{ac} at frequency f can be imaged simultaneously with the thermal signal T_{ac} that will be present at frequency $2f$, as shown in Supplementary Fig. 11.

We used the tSOT for nanoscale thermal imaging of quantum matter. Figure 2a, b shows T_{ac} images of two single-walled carbon nanotubes (CNTs) carrying an alternating current of few nA. Each CNT is wound into a loop²⁵, as outlined by the dotted trajectories and shown by SEM images in Fig. 2d, e. The thermal signal in Fig. 2a tracks the CNT, revealing the current-driven dissipation along the entire length of the CNT. Surprisingly, and in contrast to the above, Fig. 2b shows an absence of heating in the circular part of the loop. The T_{ac} image thus reveals that the applied current bypasses the loop, exposing an electrically shorted junction between the two crossing sections of the CNT. This observation illustrates the capacity of the tSOT for fault detection in operating nanodevices.

A striking feature evident in Fig. 2a, b is ring-like fine structure: a zoomed-in T_{ac} image of one of these is shown in Fig. 2c. These features resemble the Coulomb blockade rings observed in scanning gate microscopy²⁶, in which the conductance through a quantum dot is measured as a function of the position of a conducting AFM tip. In the

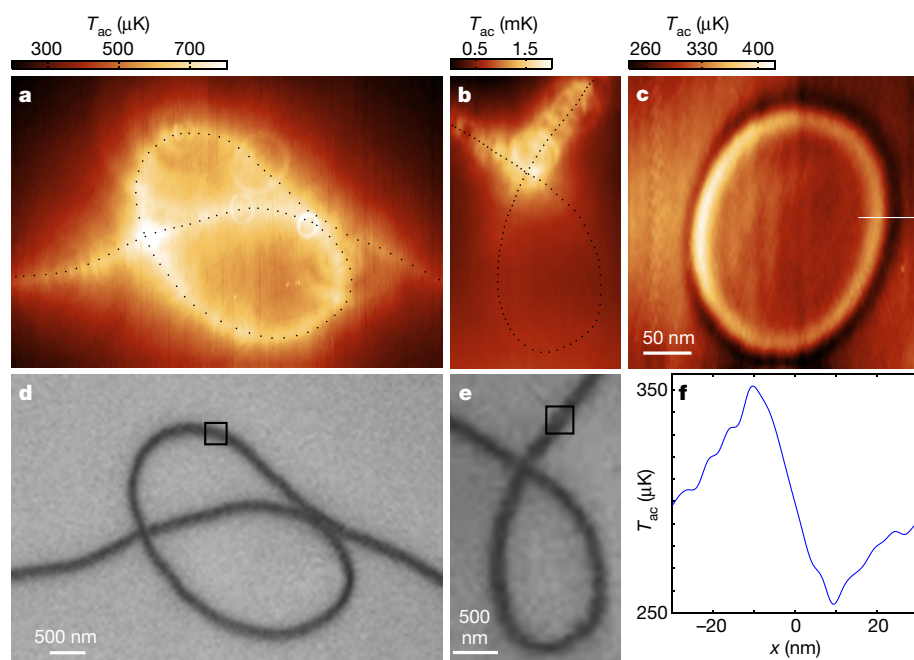


Figure 2 | Thermal imaging of single-walled CNTs and scanning gate thermometry of quantum dots. **a, b,** Thermal images of two CNT devices with loop geometry carrying I_{ac} of 12 nA (**a**) and 3 nA (**b**), revealing an electrical short at the loop intersection point in **b**. The colour scale used for T_{ac} is shown at the top of each image. The $T_{ac}(x, y)$ was acquired by a tSOT of 104-nm diameter at scanning heights of about 65 nm (**a**) and 150 nm (**b**). The ring-like structures in T_{ac} result from variations in dissipation due to modulations in the single-electron charging of individual quantum dots as described for **c**. **c,** Zoomed-in scanning gate thermometry image of a single quantum dot in the area boxed in **d** at a scan height of about 35 nm (T_{ac} scale at top). Variations in the electrochemical potential of the quantum dot induced by the scanning tSOT give rise to changes of the temperature of the dot at the Coulomb blockade peak conditions along the equipotential ring-like contour. **d, e,** SEM images of the devices in **a** and **b**. Boxed area in **e** is shown in Supplementary Fig. 15. **f,** Line-cut across the line marked on right in **c**.

case of scanning gate microscopy, the characteristic equipotential rings (corresponding to the periodic conditions of Coulomb blockade peaks governed by single-electron charging of the quantum dot²⁷) originate from the tip acting as a local gate. Since the tSOT is conducting, it can also serve as a nanoscale scanning gate. The disordered substrate potential in our long CNTs on the SiO₂/Si substrate gives rise, however, to carrier localization and formation of a series of electronic quantum dots, resulting in CNT resistances in excess of 10 MΩ. As a result, no detectable change in CNT conductance is observed while scanning the tSOT. The unique feature of the tSOT, however, is that in addition to functioning as a scanning gate it simultaneously operates as a nanoscale thermometer that detects minute changes in the local dissipation resulting from its own controllable local gating. The T_{ac} ring in Fig. 2c with a cross-sectional width of 20 nm (Fig. 2f) thus reveals microkelvin-range changes of the temperature of a quantum dot resulting from the modulation of its single-electron Coulomb blockade conductance peaks by the scanning tSOT (see Supplementary Information sections S13, S14 for additional details). This novel ‘scanning gate thermometry’ thus adds another functionality to the tSOT, allowing nanoscale manipulation of the potential and study of the induced changes in the local scattering processes and dissipation that are inaccessible by other methods.

One of the topics of great interest in the study of electronic transport at low temperatures is the non-equilibrium heating of the electron bath and the proliferation of hot carriers due to poor electron–phonon

coupling. Graphene offers a unique system in which the typical distance for such carrier–lattice cooling can exceed the dimensions of the device²⁸ owing to the anomalously long lifetime of the hot carriers²⁹. Electron cooling rates are believed to be sharply enhanced in the presence of disorder, but the precise mechanism of such enhancement is poorly understood³⁰.

To probe microscopic dissipation mechanisms in graphene, we performed scanning thermometry in high-mobility graphene encapsulated in hexagonal boron nitride (hBN). We used a washer-shaped device to which a direct current was applied between top and bottom constrictions, as illustrated in Fig. 3a (see Supplementary Information section S12 for details). Figure 3b shows the corresponding scanning gate thermal image, revealing a startling visualization of dissipation processes, which are manifested in a complex structure of two ‘necklaces’ of rings along the inner and outer edges of the device. The sharp ring-like structures have the same origin as the ring-like patterns in the CNTs in Fig. 2a–c, and reveal the presence of localized resonant states at the edges of the graphene; these states are acting as atomic-scale heat beacons. The tSOT functions as a top gate, which tunes the potential of the localized states into resonance when the combination of its distance and the voltage V_{tSOT} applied between the tip and the sample match the resonance conditions. This gives rise to sharp rings of enhanced temperature, as described in detail in Supplementary Information sections S13 and S15. Consequently, the ring radius R changes

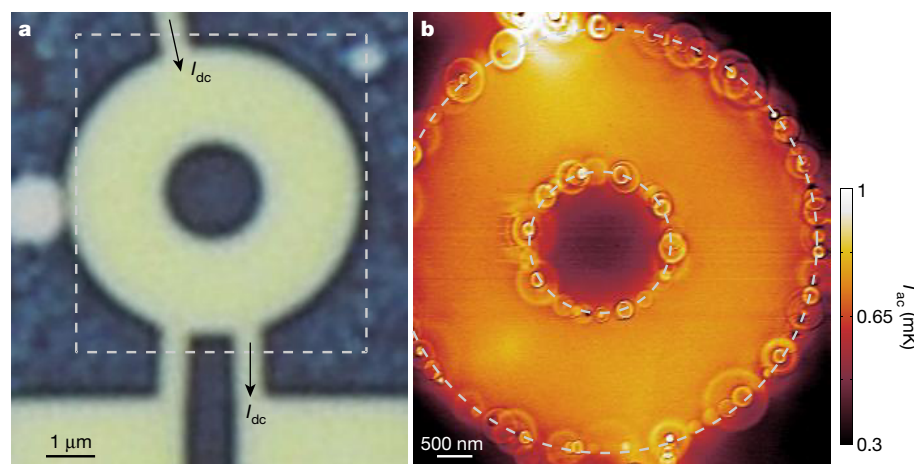


Figure 3 | Scanning gate thermometry of dissipation at localized resonant states at graphene edges. **a,** Optical image of hBN/graphene/hBN structure patterned into a washer shape (bright) with inner and outer diameters of 2 μm and 6 μm, respectively. Arrows show where the direct current I_{dc} was applied. **b,** Scanning a.c. gate thermometry T_{ac} image of the boxed area in **a** in the presence of $I_{dc} = 6 \mu A$ at a carrier concentration of 10^{12} cm^{-2} ; the T_{ac} colour scale is shown on right. The $T_{ac}(x, y)$ was acquired by a tSOT of 100-nm diameter at 4.2 K at a height of 40 nm in the presence of an a.c. potential of 200 mV applied to the tip (see Supplementary Information sections S13–S15). The dashed lines mark graphene edges. The ‘necklace’ of rings reveals the presence of resonant states along the edges of graphene that serve as local centres of energy dissipation.

on varying V_{tSOT} , as demonstrated in Supplementary Fig. 16 and Supplementary Video S1, or on changing the tip height, as demonstrated in the case of a CNT in Supplementary Fig. 15.

Despite the similarity between the ring-like structures observed in CNTs and in graphene, their microscopic origin appears to be different. For CNTs, the disorder potential creates quantum dots that confine discrete electronic states. Since the conductance of the quantum dot is sharply enhanced under resonant conditions, the Coulomb blockade staircase gives rise to the sequence of concentric rings of enhanced local temperature in the presence of current flow through the CNT (Supplementary Fig. 15). In our hBN/graphene/hBN devices, in contrast, the hot electrons flow mainly in the bulk of the graphene, while electron–lattice cooling occurs predominantly at defect states at graphene edges. Such defects can be formed by the vacancies and adatoms produced at the exposed graphene edges during the etching process. Spatially localized electronic states originating from such defects, with energies pinned to the Dirac point³¹, were recently revealed by scanning tunnelling microscopy³². Localized states can considerably enhance electron–phonon coupling, thus providing a local drain for the excess energy of the hot electrons³⁰. Since each defect creates a single resonant electronic state³¹, only one ring should be observed around each defect with radius R that depends on V_{tSOT} , as shown in Supplementary Fig. 16 and Supplementary Video S1. Acting as gate-tunable spatially localized energy flow bottlenecks, these dissipation centres are likely to play a dominant role in the hot-carrier applications of graphene electronics. These observations shed new light on electron–lattice cooling mechanisms in graphene.

The observation of striking, spatially localized dissipation centres at graphene edges underscores the potential of the tSOT technique for uncovering the microscopic origins of dissipation in novel states of matter. Other systems of interest are topologically protected surface states, edge states in quantum spin and anomalous quantum Hall systems, and surface states in Weyl semimetals. By choosing appropriate superconducting materials^{22,23,33}, it should be possible to extend the operating temperature range of the tSOT from tens of millikelvin to tens of kelvin (Supplementary Fig. 1d), thus allowing investigation of a wide range of quantum systems. In addition, the operation of the tSOT at elevated magnetic fields combined with the multifunctional abilities of magnetic sensing and scanning gate thermometry opens the way to nanoscale investigation and imaging of intricate thermoelectric and thermomagnetic phenomena. These include the Nernst effect, the thermal Hall effect, thermoelectric nanoscale cooling and quantum heat conductance.

Received 25 November 2015; accepted 31 August 2016.

Published online 26 October 2016.

1. Yue, Y. & Wang, X. Nanoscale thermal probing. *Nano Rev.* **3**, 11586 (2012).
2. Giazotto, F., Heikkilä, T. T., Luukanen, A. M. & Pekola, J. P. Opportunities for mesoscopies in thermometry and refrigeration: physics and applications. *Rev. Mod. Phys.* **78**, 217–274 (2006).
3. Jin, C. Y., Li, Z., Williams, R. S., Lee, K. C. & Park, I. Localized temperature and chemical reaction control in nanoscale space by nanowire array. *Nano Lett.* **11**, 4818–4825 (2011).
4. Kucsko, G. *et al.* Nanometre-scale thermometry in a living cell. *Nature* **500**, 54–58 (2013).
5. Mecklenburg, M. *et al.* Nanoscale temperature mapping in operating microelectronic devices. *Science* **347**, 629–632 (2015).
6. Teyssieux, D., Thiery, L. & Cretin, B. Near-infrared thermography using a charge-coupled device camera: application to microsystems. *Rev. Sci. Instrum.* **78**, 034902 (2007).
7. Toyli, D. M., de las Casas, C. F., Christle, D. J., Dobrovitski, V. V. & Awschalom, D. D. Fluorescence thermometry enhanced by the quantum coherence of single spins in diamond. *Proc. Natl Acad. Sci. USA* **110**, 8417–8421 (2013).
8. Neumann, P. *et al.* High-precision nanoscale temperature sensing using single defects in diamond. *Nano Lett.* **13**, 2738–2742 (2013).
9. Reparaz, J. S. *et al.* A novel contactless technique for thermal field mapping and thermal conductivity determination: two-laser Raman thermometry. *Rev. Sci. Instrum.* **85**, 034901 (2014).
10. Brites, C. D. S. *et al.* Thermometry at the nanoscale. *Nanoscale* **4**, 4799–4829 (2012).

11. Majumdar, A. Scanning thermal microscopy. *Annu. Rev. Mater. Sci.* **29**, 505–585 (1999).
12. Menges, F. *et al.* Temperature mapping of operating nanoscale devices by scanning probe thermometry. *Nat. Commun.* **7**, 10874 (2016).
13. Kim, K., Jeong, W., Lee, W. & Reddy, P. Ultra-high vacuum scanning thermal microscopy for nanometer resolution quantitative thermometry. *ACS Nano* **6**, 4248–4257 (2012).
14. Sadat, S., Tan, A., Chua, Y. J. & Reddy, P. Nanoscale thermometry using point contact thermocouples. *Nano Lett.* **10**, 2613–2617 (2010).
15. Grosse, K. L., Bae, M. H., Lian, F., Pop, E. & King, W. P. Nanoscale Joule heating, Peltier cooling and current crowding at graphene–metal contacts. *Nat. Nanotechnol.* **6**, 287–290 (2011).
16. Pekola, J. P. Towards quantum thermodynamics in electronic circuits. *Nat. Phys.* **11**, 118–123 (2015).
17. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.* **5**, 183–191 (1961).
18. Bérut, A. *et al.* Experimental verification of Landauer’s principle linking information and thermodynamics. *Nature* **483**, 187–189 (2012).
19. Herr, Q. P., Herr, A. Y., Oberg, O. T. & Ioannidis, A. G. Ultra-low-power superconductor logic. *J. Appl. Phys.* **109**, 103903 (2011).
20. Kuhn, K. J. *et al.* The ultimate CMOS device and beyond. *IEEE Int. Electron Devices Meet.* 8.1.1–8.1.4 (2012).
21. Faivre, T., Golubev, D. & Pekola, J. P. Josephson junction based thermometer and its application in bolometry. *J. Appl. Phys.* **116**, 094302 (2014).
22. Vasyukov, D. *et al.* A scanning superconducting quantum interference device with single electron spin sensitivity. *Nat. Nanotechnol.* **8**, 639–644 (2013).
23. Finkler, A. *et al.* Self-aligned nanoscale SQUID on a tip. *Nano Lett.* **10**, 1046–1049 (2010).
24. Schwab, K., Henriksen, E. A., Worlock, J. M. & Roukes, M. L. Measurement of the quantum of thermal conductance. *Nature* **404**, 974–977 (2000).
25. Shadmi, N., Geblinger, N., Ismach, A. & Joselevich, E. Formation of ordered vs disordered carbon nanotube serpentine on anisotropic vs isotropic substrates. *J. Phys. Chem. C* **118**, 14044–14050 (2014).
26. Woodside, M. T. & McEuen, P. L. Scanned probe imaging of charge states in nanotube quantum dots. *Science* **296**, 1098–1101 (2002).
27. Grabert, H. & Devoret, M. H. (eds) *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures* (Vol. 294, NATO Sci. Ser. B, Springer, 2013).
28. Bandurin, D. A. *et al.* Negative local resistance caused by viscous electron backflow in graphene. *Science* **351**, 1055–1058 (2016).
29. Bistritzer, R. & MacDonald, A. H. Electronic cooling in graphene. *Phys. Rev. Lett.* **102**, 206410 (2009).
30. Song, J. C. W., Reizer, M. Y. & Levitov, L. S. Disorder-assisted electron-phonon scattering and cooling pathways in graphene. *Phys. Rev. Lett.* **109**, 106602 (2012).
31. Pereira, V. M., Guinea, F., Lopes dos Santos, J. M. B., Peres, N. M. R. & Castro Neto, A. H. Disorder induced localized states in graphene. *Phys. Rev. Lett.* **96**, 036801 (2006).
32. González-Herrero, H. *et al.* Atomic-scale control of graphene magnetism by using hydrogen atoms. *Science* **352**, 437–441 (2016).
33. Lachman, E. *et al.* Visualization of superparamagnetic dynamics in magnetic topological insulators. *Sci. Adv.* **1**, e1500740 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. F. Young for discussions, M. V. Costache and S. O. Valenzuela for facilitation of fabrication of permalloy (Py)/copper (Cu) samples that were used in Supplementary Information section S9, D. Shahar, I. Tamir, T. Levinson and S. Mitra for assistance in fabrication of a:In₂O₃ integrated devices that were used in Supplementary Information section S2, M. E. Huber for SOT readout setup, and M. L. Rappaport for technical assistance. This work was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 programme (grant no. 655416), by the Minerva Foundation with funding from the Federal German Ministry of Education and Research, and by a Rosa and Emilio Segré Research Award. L.S.L. and E.Z. acknowledge the support of the MISTI MIT-Israel Seed Fund.

Author Contributions D.H., J.C. and E.Z. conceived the technique and designed the experiments. D.H. and J.C. performed the measurements. D.H. performed the analysis and theoretical modelling. L.E. constructed the scanning SOT microscope. M.B.S. and A.K.G. designed and provided the graphene sample and contributed to the analyses of the results. N.S. and E.J. fabricated the CNT samples. J.C. fabricated the Cu/Py sample. D.H., H.R.N. and J.S. fabricated the a:In₂O₃ sample. D.H. and Y.R. designed and fabricated the spatial resolution demonstration sample. H.R.N., Y.A. and A.U. fabricated the tSOT sensors. Y.A. and Y.M. developed the SOT fabrication technique. A.U., Y.M. and D.H. developed the tuning-fork based tSOT height control technique. L.S.L. performed theoretical analysis. D.H., J.C. and E.Z. wrote the manuscript. All authors participated in discussions and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H. (dorri.halbental@weizmann.ac.il) or E.Z. (eli.zeldov@weizmann.ac.il).

Reviewer Information Nature thanks I. Maasilta, S. Volz and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Intrinsically stretchable and healable semiconducting polymer for organic transistors

Jin Young Oh^{1*}, Simon Rondeau-Gagné^{1*†}, Yu-Cheng Chiu^{1*†}, Alex Chortos¹, Franziska Lissel¹, Ging-Ji Nathan Wang¹, Bob C. Schroeder^{1†}, Tadanori Kurosawa¹, Jeffrey Lopez¹, Toru Katsumata^{1,2}, Jie Xu¹, Chenxin Zhu³, Xiaodan Gu^{1,4}, Won-Gyu Bae¹, Yeongin Kim³, Lihua Jin^{5†}, Jong Won Chung^{1,6}, Jeffrey B.-H. Tok¹ & Zhenan Bao¹

Thin-film field-effect transistors are essential elements of stretchable electronic devices for wearable electronics^{1,2}. All of the materials and components of such transistors need to be stretchable and mechanically robust^{3,4}. Although there has been recent progress towards stretchable conductors^{5–8}, the realization of stretchable semiconductors has focused mainly on strain-accommodating engineering of materials, or blending of nanofibres or nanowires into elastomers^{9–11}. An alternative approach relies on using semiconductors that are intrinsically stretchable, so that they can be fabricated using standard processing methods¹². Molecular stretchability can be enhanced when conjugated polymers, containing modified side-chains and segmented backbones, are infused with more flexible molecular building blocks^{13,14}. Here we present a design concept for stretchable semiconducting polymers, which involves introducing chemical moieties to promote dynamic non-covalent crosslinking of the conjugated polymers. These non-covalent crosslinking moieties are able to undergo an energy dissipation mechanism through breakage of bonds when strain is applied, while retaining high charge transport abilities. As a result, our polymer is able to recover its high field-effect mobility performance (more than 1 square centimetre per volt per second) even after a hundred cycles at 100 per cent applied strain. Organic thin-film field-effect transistors fabricated from these materials exhibited mobility as high as 1.3 square centimetres per volt per second and a high on/off current ratio exceeding a million. The field-effect mobility remained as high as 1.12 square centimetres per volt per second at 100 per cent strain along the direction perpendicular to the strain. The field-effect mobility of damaged devices can be almost fully recovered after a solvent and thermal healing treatment. Finally, we successfully fabricated a skin-inspired stretchable organic transistor operating under deformations that might be expected in a wearable device.

Incorporation of dynamic non-covalent crosslinking between flexible polymer chains is an important method of achieving high stretchability and self-healing properties^{15,16}. The dynamic bonds can easily be broken to allow energy dissipation upon strain, thus making the system more tolerant of strain and mechanical stimuli. Moreover, the bonds can be reformed to recover the initial mechanical property and self-heal. Among the different types of dynamic bonds, hydrogen bonds are particularly suitable for skin-inspired electronics owing to their spontaneous formation and healing ability^{17–20}. Such polar substituents in organic semiconductors have received increased attention in recent organic electronics research, but mostly in producing self-assembled nanostructures instead of using them to improve mechanical properties²¹.

2,6-pyridine dicarboxamide (PDCA) was chosen to introduce hydrogen bonding within the flexible polymer backbone since this unit contains two amide groups possessing moderate hydrogen-bonding strength, allowing the formation of a polymer network without drastically increasing the tensile modulus of the material^{22,23}. Previous work has indicated that introducing a small fraction of non-conjugated units to the polymer backbone does not noticeably degrade the charge transport mobility²⁴. Here we introduce alkyl spacers to enhance the flexibility of the dynamic moieties. Semiconducting polymers incorporating different ratios of non-conjugated PDCA moiety were thus synthesized (**P1** to **P4**; structures shown in Fig. 1a). To confirm the presence of hydrogen bonds, both X-ray crystallography and nuclear magnetic resonance (NMR) experiments (Extended Data Figs 1 and 2, and Supplementary Figs 1, 2 and 11) were used to ascertain hydrogen-bond formation in PDCA-containing model compounds.

Initial evaluation of the electrical properties was performed using organic thin-film field-effect transistor (OTFT) devices (Fig. 1b). For devices fabricated with **P2** and **P3**, the measured mobilities were both $>1\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$. Even with **P4** (containing 20 mol% of non-conjugated monomers), its field-effect mobility remained as high as $0.58\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ (Supplementary Fig. 12). To understand further the contributions from the disruption of conjugation and incorporation of hydrogen bonding on the mechanical properties of the polymers, we next measured the elastic modulus of the polymers (Fig. 1c and Supplementary Fig. 13). All the polymers with PDCA moieties exhibited a lower elastic modulus than that of the fully conjugated polymer (**P1**). Although intermolecular hydrogen bonding essentially cross-links the polymers and thus should increase the elastic modulus of the polymer film²⁵, it appears that disruption of conjugation—that is, reducing the rigidity of the conjugated polymer backbone—has a larger effect on the elastic modulus of the polymer semiconductor film. We attributed the reduction in the measured elastic modulus to the changes in film morphology, such as the increase in the amorphous fraction in the polymer film or the decrease in relative crystallinity (Fig. 1e), and also a slight reduction of the average size of crystallites (Extended Data Fig. 3c), as characterized by grazing incidence X-ray diffraction analysis (Supplementary Fig. 14).

To gain further insight into the molecular level changes during stretching, the degree of polymer chain alignment under strain was measured using polarized ultraviolet–visible spectroscopy and quantified using dichroic ratios (Fig. 1d and Supplementary Fig. 16). The dichroic ratio of **P1** initially increased slightly with strain owing to strain-induced chain alignment, but soon decreased to approximately 1; this was caused by chain relaxation due to crack formation upon

¹Department of Chemical Engineering, Stanford University, Stanford, California 94305-5025, USA. ²Corporate Research and Development, Performance Materials Technology Center, Asahi Kasei Corporation, 2-1 Samejima, Fuji, Shizuoka 416-8501, Japan. ³Department of Electrical Engineering, Stanford University, Stanford, California 94305-5025, USA. ⁴Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ⁵Department of Civil and Environmental Engineering, Stanford University, Stanford, California 94305-5025, USA. ⁶Samsung Advanced Institute of Technology, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-803, South Korea. [†]Present addresses: Department of Chemistry and Biochemistry, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4, Canada (S.R.-G.); Department of Chemical Engineering and Materials Science, Yuan Ze University, Taoyuan 32003, Taiwan (Y.-C.C.); Materials Research Institute and School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK (B.C.S.); Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, Los Angeles, California 90095, USA (L.J.).

*These authors contributed equally to this work.

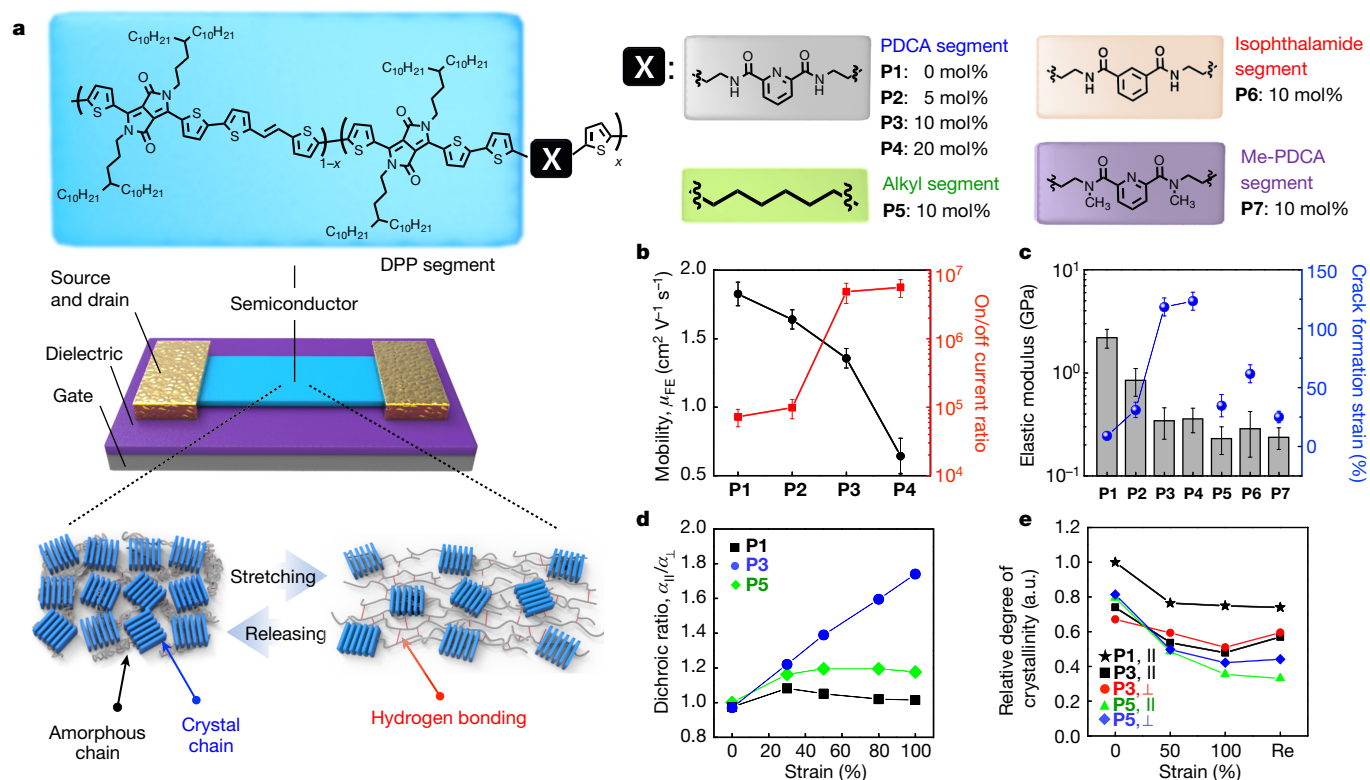


Figure 1 | Design of stretchable and healable semiconducting polymer OTFTs and their performance characterizations. **a**, Chemical structures of polymers **P1** to **P7** and general mechanism for enhancement of stretchability in conjugated polymers via dynamic bonding. The stretchable semiconducting polymers are based on 3,6-di(thiophen-2-yl)-2,5-dihydropyrrolo[3,4-c]pyrrole-1,4-dione (DPP) repeating units and non-conjugated 2,6-pyridine dicarboxamide (PDCA) moieties introduced directly in the polymer backbone as hydrogen bonding units. To qualitatively evaluate the stretching properties of the polymers, thin films are supported on polydimethylsiloxane (PDMS) substrates and stretched to different strain ratios (0% to 100% strain; 'Re' refers to the value after strain release). **b**, Field-effect mobility μ_{FE} and on/off current ratios of **P1** to **P4** as measured by conventional OTFTs (the source and drain electrodes are 40-nm-thick Au, the dielectric layer is 300-nm-thick SiO₂ and the gate

electrode is highly doped silicon substrate). **c**, Influence of the hydrogen-bonding moiety amount on elastic modulus and crack formation strain. Error bars in **b** (three batches of samples were prepared, and five OTFTs from each sample were measured) and **c** (one data point was obtained from each sample, and three batches of samples were prepared and measured) are mean \pm s.d. **d**, Dichroic ratios of **P1**, **P3** and **P5** under various strain levels. $\alpha_{||}$ and α_{\perp} are absorption intensities measured with the polarization direction of light parallel and perpendicular to the stretching direction, respectively. **e**, Influence of the strain ratio on the relative degree of crystallinity extracted from peak (200) for both 'parallel' and 'perpendicular' directions of **P1**, **P3** and **P5**. Owing to the formation of large cracks, the relative degree of crystallinity of **P1** perpendicular to the strain direction was not measured. The semiconductor is 35 nm thick. (a.u., arbitrary units.)

increasing the strain from 30% to 100%. This observation indicates that **P1** is not very stretchable and, indeed, we observed that cracks began to form even at low strains (5%–10%); these were further confirmed by both optical microscopy and atomic force microscopy (AFM). On the other hand, the dichroic ratio of **P3** increased linearly to 1.8, as strain increased from 30% to 100%. This observation is consistent with our microscopic observations that no microscale cracks were formed for **P3**, even with strain as high as 100%, showing that polymer chain alignment by strain can be achieved (Supplementary Figs 17 and 18 and Extended Data Fig. 3).

Relative degree of crystallinity analysis was used next to examine the change of morphology under strain (Fig. 1e and Supplementary Figs 19 and 20). The relative degree of crystallinity of **P1** decreased with strain from 0% to 50%, and plateaued from 50% onwards. This observation, combined with AFM studies, indicates that the tensile strain applied to the thin film was initially released through breakage of the crystalline regions until crack formation. In contrast, the relative degree of crystallinity of **P3** decreased steadily upon applied strain from 0% to 100%. Upon releasing the strain, we observed that the relative degree of crystallinity recovered only partially, indicating that some of its crystalline domains may be permanently altered or broken apart by the applied strain. This result, combined with the steady increase in dichroic ratio previously observed, suggests that **P3** has multiple mechanisms for energy dissipation during strain: (1) stretching and

realignment of polymer chains in amorphous regions; (2) breaking of some of the crystalline domains; and (3) breakage of hydrogen bonding.

The mechanism of energy dissipation through hydrogen-bond breakages has previously been observed¹⁸. We confirmed this benefit of incorporating hydrogen-bonding sites as strain-releasing moieties by comparison with several reference polymers. **P5** contains 10 mol% of non-conjugated alkyl spacers, that is, no PDCA moieties, on its polymer backbone (Fig. 1a). Despite having a lower elastic modulus than **P3**, **P5** can achieve a maximum strain of only 30% before cracks form, a value much lower than that of **P3**'s maximum strain (110%). Furthermore, the measured maximum dichroic ratio for **P5** was found to be only about 1.16 at 30% strain, consistent with the strain value at the onset of crack formation. In addition, the relative crystallinity of **P5** was measured to be lower than **P1**, indicating that an increase in the fraction of amorphous regions has occurred within the polymer thin film, similar to **P2**–**P4**.

However, when 100% strain was applied to the film, the relative degree of crystallinity decreased drastically from 0.80 to 0.35, indicating that the tensile energy in **P5** film is dissipated primarily through the breaking of crystallites (Supplementary Fig. 21). This observation is consistent with a recent report, which did not show much improvement on fracture strain with conjugation breakage of another DPP polymer²⁴. **P6** is another reference polymer, in which the pyridine moiety was replaced by a benzene ring while the amide moieties remain (Fig. 1a).

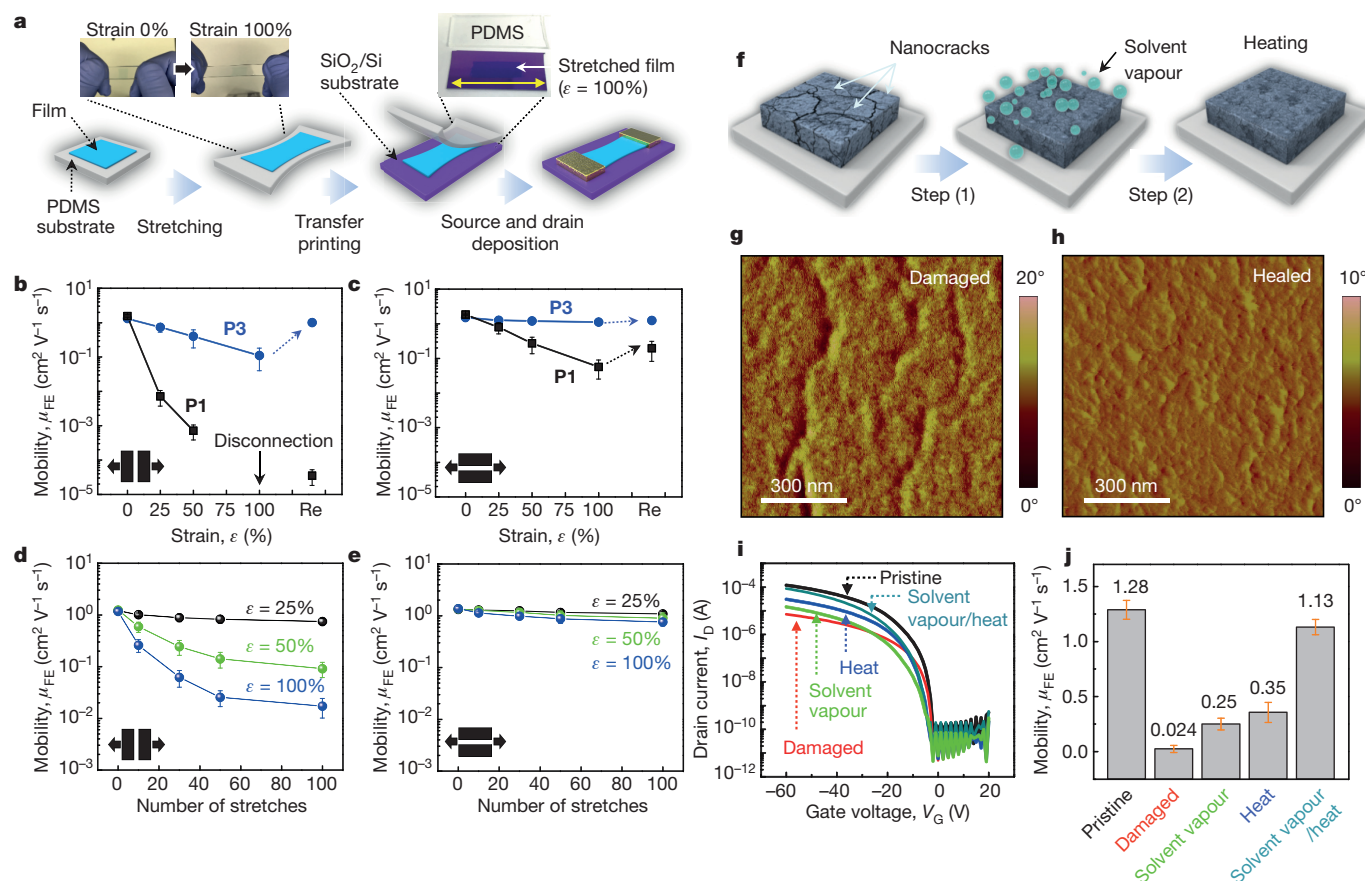


Figure 2 | Charge transport of semiconducting polymers under applied strain and their healing properties. **a**, Fabrication process for OTFTs with stretched polymer thin films. **b**, **c**, Field-effect mobility μ_{FE} as a function of various strains along the strain direction (**b**) and perpendicular to strain direction (**c**). **d**, **e**, Field-effect mobility versus number of stretching cycles performed along the strain direction (**d**) and perpendicular to

P6 was observed to possess enhanced stretchability compared to polymers **P1** and **P5** (Fig. 1c and Supplementary Fig. 18f), which again suggests the importance of the intermolecular hydrogen bonding for enhanced mechanical properties. However, as compared to polymers **P2**, **P3** and **P4** (all containing pyridine moieties), its stretchability was reduced. This suggests that the introduction of pyridine moieties may contribute to the enhanced mechanical properties either through participating in intra- and intermolecular hydrogen-bonding or through changing the morphology of the film. Finally, reference polymer **P7** was synthesized to remove the hydrogen-bonding amide hydrogens (Fig. 1a). **P7** showed a decreased stretchability with an onset strain for crack formation of about 25% (Fig. 1c and Supplementary Fig. 18g). The above control experiments with reference polymers further confirm the important contribution of the dynamic non-covalent hydrogen-bond-forming moieties as strain-releasing elements for enhancing the stretchability of conjugated polymers.

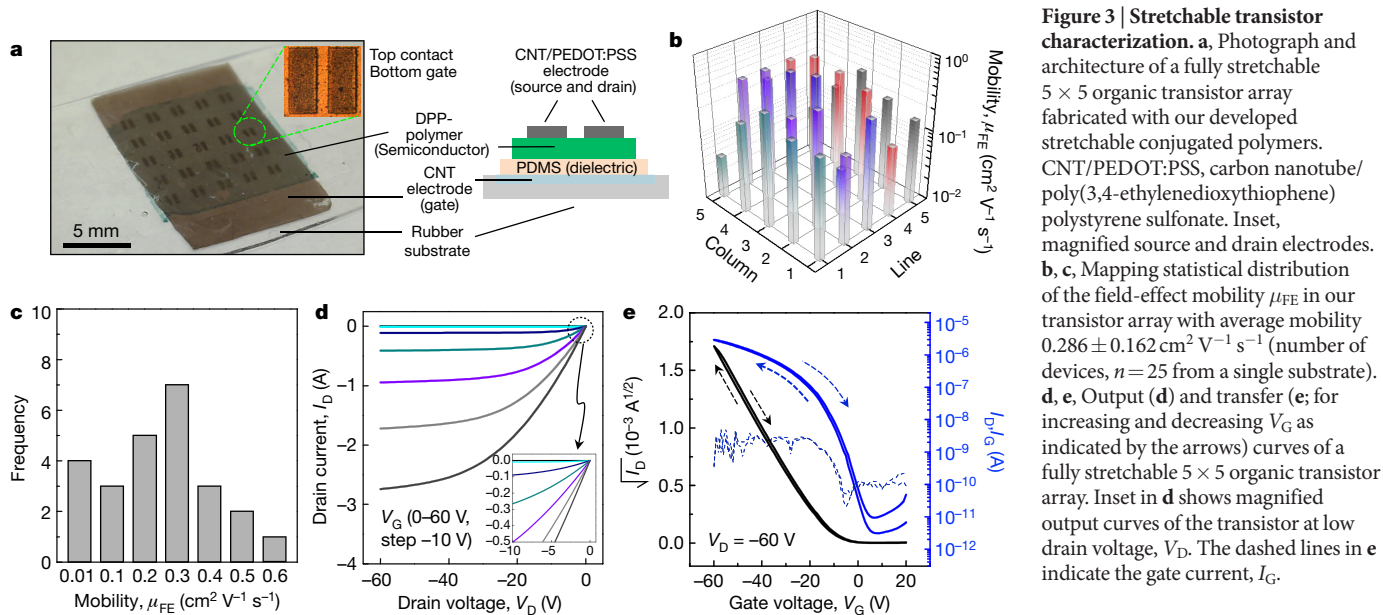
To evaluate the charge transport behaviour of our fabricated stretchable semiconducting polymers, OTFTs were fabricated and characterized (Fig. 2a). On applying strains up to 100%, we observed that the average field-effect mobility of the **P3** device decreased moderately, from $1.32 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ to $0.11 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$, along the direction of applied strain. Furthermore, the mobility was observed to recover to $1.00 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ upon releasing the applied strain. When the strain is applied in a perpendicular direction, the mobility of the **P3** device is maintained, that is, $>1 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$, up to 100% strain and even after release. In contrast, the mobility of the **P1** device degraded substantially even at 25% applied strain, and was totally non-functional at 100% strain. Unlike **P3**, the mobility of the **P1** device is unable to recover upon stress release.

strain direction (**e**). **f**, Representation of the treatments used for healing the conjugated polymer films. **g**, **h**, AFM phase image for damaged (**g**) and healed (**h**) film of **P3**. **i**, **j**, Transfer curves (**i**) and field-effect mobility (**j**) of damaged and healed **P3** OTFTs. Error bars in **b**, **c**, **d**, **e** and **j** are mean \pm s.d. (three batches of samples were prepared, and three OTFTs from each sample were measured).

Besides a single stretching event, we also performed rigorous repeated stretching cycle tests on the **P3** devices at various strains (Fig. 2d, e). After stretching the device for 100 cycles between 0% and 25% strain, the mobility along the stretching direction was decreased by about 40% (that is, from $1.2 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ to $0.74 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$) and was further reduced to $0.017 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ when subjected to 100% strain. However, when the strain was along the perpendicular direction, the device showed higher durability and robustness. Even after 100 cycles up to 100% strain, only a 26% decrease in mobility was observed. Only a 19% decrease in mobility was observed for 100 cycles up to 25% strain (Supplementary Fig. 22).

In addition to intrinsic stretchability, the healing ability enabled by the dynamic bonding represents another major advantage of the polymer design. Although the self-healing abilities of non-conjugated polymer networks cross-linked with hydrogen-bonding sites have been reported²⁶, self-healing ability has not been investigated for conjugated polymers. To facilitate efficient healing for damaged **P3** films, post treatments via heat and solvent annealing are required to promote polymer chain movement (Fig. 2f). When applied independently, both thermal and solvent annealing greatly reduced the size and density of the nanocracks. However, the field-effect mobility was observed to be only slightly recovered (Fig. 2g–j and Supplementary Fig. 23).

Solvent annealing applied in conjunction with thermal annealing promoted the most efficient healing of the polymer films. Specifically, we observed a complete disappearance of the nanocracks within the damaged films, and more importantly, an almost complete recovery of the average field-effect mobility to $1.13 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$ (Fig. 2i, j and Supplementary Fig. 23). Additionally, the dependence of dichroic ratio



on strain completely recovered to a level similar to that of a pristine undamaged film, indicating the healed film has most probably returned to its original film morphology. In comparison, our various applied healing conditions did not improve the damaged **P1** and **P5** films (Supplementary Figs 24–28 and Extended Data Fig. 4). To the best

of our knowledge, this is the first demonstration of complete healing (albeit non-autonomous) of a high field-effect mobility conjugated polymer after mechanical damages.

Equipped with our newly developed high mobility and high stretchability polymer semiconductor, we proceed to fabricate fully stretchable

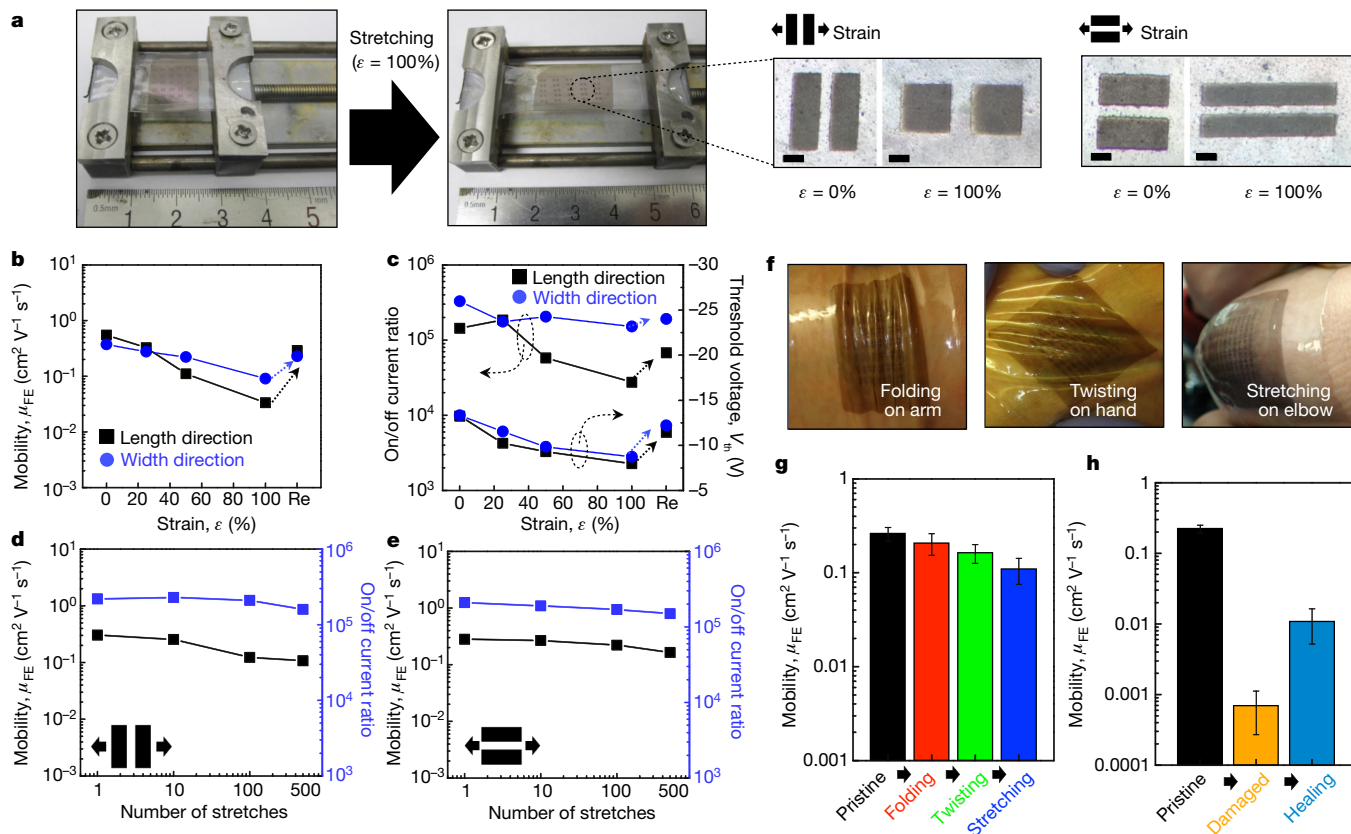


Figure 4 | Stretchable transistor as skin-inspired wearable device. **a**, *In situ* stretching images of stretchable OTFT from 0% to 100% strain (scale bars in close-ups, $300 \mu\text{m}$). **b**, **c**, Field-effect mobility μ_{FE} (**b**) and on/off current ratio and threshold voltage V_{th} (**c**) of a stretchable transistor during a stretching cycle. **d**, **e**, Mobility and on/off ratio of OTFTs for 500 stretching cycles at 25% strain along channel length and width directions,

respectively. **f**, Photographs of folded, twisted and stretched OTFTs on human skin. **g**, OTFT mobilities after various extreme human motions. **h**, Mobilities of mechanically damaged OTFTs before and after healing process. Error bars in **g** and **h** are mean \pm s.d. (five batches of samples were prepared, and one OTFT from each sample was measured).

OTFTs (5×5 arrays; Fig. 3a, Supplementary Fig. 29 and Extended Data Fig. 5). Our obtained device yield is 100%, along with good switching performance (Fig. 3b–e). Furthermore, most of the devices exhibited field-effect mobilities in the range of about $10^{-1} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ with $>10^5$ on/off current ratio (Supplementary Fig. 29). In specific, the highest value obtained for the mobility is $0.6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, while the average mobility was $0.28 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. In general, we observed that the fully stretchable OTFTs all exhibited lower mobilities compared to OTFTs as fabricated on rigid Si/SiO₂ substrates treated with self-assembled monolayer of *n*-octadecyltrimethoxysilan (OTS). We attributed this observation to the lower applied electric field with a thicker layer of dielectric on the stretchable OTFTs devices (Supplementary Fig. 30). The stretchable devices showed little hysteresis and low gate-leakage currents. These measured performance parameters represent the best of the current reported stretchable organic semiconductors and fully stretchable OTFTs (Supplementary Tables 3 and 4)^{27–30}.

To verify the stretchability of our organic transistors, we investigated their electrical performance by subjecting them to various strenuous mechanical strain conditions. The transistors showed a slow linear decrease in mobility when being strained up to 100%. After releasing the strain, the mobility was observed to recover to close to its initial values (Fig. 4b and Supplementary Fig. 31). The same trend was also observed for the on/off current ratios. These trends are qualitatively similar to the trend observed when only the polymer semiconductor is stretched. Most practical applications, however, require effective operation only for applied strains of 20% to 30%, so we performed stretching durability tests at 25% strain and 500 cycles for our fabricated stretchable transistors (Fig. 4d, e). We observed that both the field-effect mobility and on/off current ratio of our transistors did not suffer any noticeable decrease in performance. Furthermore, we mounted our fabricated devices on human limbs to test device tolerance to various common movements, such as folding of the arm, twisting of the hand, and stretching of the elbow. Our device is able to maintain its average mobility at $>0.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (Supplementary Fig. 32) under all these conditions. Finally, the healing ability of the polymer semiconductor can again be observed using the heating and solvent annealing process we developed (Fig. 4h and Supplementary Fig. 33).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 February; accepted 12 September 2016.

- Chortos, A. & Bao, Z. Skin-inspired electronic devices. *Mater. Today* **17**, 321–331 (2014).
- Wagner, S. & Bauer, S. Materials for stretchable electronics. *MRS Bull.* **37**, 207–213 (2012).
- Savagatrup, S. *et al.* Molecularly stretchable electronics. *Chem. Mater.* **26**, 3028–3041 (2014).
- Hammock, M. L. *et al.* 25th anniversary article: the evolution of electronic skin (e-skin): a brief history, design, consideration, and recent progress. *Adv. Mater.* **25**, 5997–6038 (2013).
- Yao, S. & Zhu, Y. Nanomaterial-enabled stretchable conductors: strategies, materials, and devices. *Adv. Mater.* **27**, 1480–1511 (2015).
- Benight, S. J., Wang, C., Tok, J. B. H. & Bao, Z. Stretchable and self-healing polymers and devices for electronic skin. *Prog. Polym. Sci.* **38**, 1961–1977 (2013).
- Oh, J. Y. *et al.* Conducting polymer dough for deformable electronics. *Adv. Mater.* **28**, 4455–4461 (2016).
- Sekitani, T. *et al.* Stretchable active-matrix organic light-emitting diode display using printable elastic conductors. *Nat. Mater.* **8**, 494–499 (2009).
- Kaltenbrunner, M. *et al.* Ultrathin and lightweight organic solar cells with high flexibility. *Nat. Commun.* **3**, 770 (2012).
- Khang, D.-Y., Jiang, H., Huang, Y. & Rogers, J. A. A stretchable form of single-crystal silicon for high performance electronics on rubber substrates. *Science* **311**, 208–212 (2006).
- Shin, M. *et al.* Polythiophene nanofibril bundles surface-embedded in elastomer: a route to a highly stretchable active channel layer. *Adv. Mater.* **27**, 1255–1261 (2015).
- O'Connor, B. *et al.* Anisotropic structure and charge transport in highly strain-aligned regioregular poly(3-hexylthiophene). *Adv. Funct. Mater.* **21**, 3697–3705 (2011).
- Müller, C. *et al.* Tough, semiconducting polyethylene-poly(3-hexylthiophene) diblock copolymers. *Adv. Funct. Mater.* **17**, 2674–2679 (2007).

- Printz, A. D. *et al.* Increased elasticity of a low-bandgap conjugated copolymer by random segmentation for mechanically robust solar cells. *RSC Adv.* **4**, 13635–13643 (2014).
- Yang, Y. & Urban, M. W. Self-healing polymeric materials. *Chem. Soc. Rev.* **42**, 7446–7467 (2013).
- Black Ramirez, A. L. *et al.* Mechanochemical strengthening of a synthetic polymer in response to typically destructive shear forces. *Nat. Chem.* **5**, 757–761 (2013).
- Chen, Y., Kushner, A. M., Williams, G. A. & Guan, Z. Multiphase design of autonomic self-healing thermoplastic elastomers. *Nat. Chem.* **4**, 467–472 (2012).
- Sun, J. Y. *et al.* Highly stretchable and tough hydrogels. *Nature* **489**, 133–136 (2012).
- Cordier, P., Tournilhac, F., Soulié-Ziakovic, C. & Leibler, L. Self-healing and thermoreversible rubber from supramolecular assembly. *Nature* **451**, 977–980 (2008).
- Yuk, H. *et al.* Tough bonding of hydrogels to diverse non-porous surfaces. *Nat. Mater.* **15**, 190–196 (2015).
- Gsänger, M. *et al.* Organic semiconductors based on dyes and color pigments. *Adv. Mater.* **28**, 3615–3645 (2016).
- Ray, M., Ghosh, D., Shirin, Z. & Mukherjee, R. Highly stabilized low-spin iron(III) and cobalt(III) complexes of a tridentate bis-amide ligand 2,6-bis (N-phenylcarbamoyl)pyridine. Novel nonmacrocyclic tetraamido-N coordination and two unusually short metal-pyridine bonds. *Inorg. Chem.* **36**, 3568–3572 (1997).
- Marlin, D. S., Olmstead, M. M. & Mascharak, P. K. Extended structures controlled by intramolecular and intermolecular hydrogen bonding: a case study with pyridine-2,6-dicarboxamide, 1,3-benzenedicarboxamide and N, N'-dimethyl-2,6-pyridinedicarboxamide. *J. Mol. Struct.* **554**, 211–223 (2000).
- Savagatrup, S. *et al.* Effect of broken conjugation on the stretchability of semiconducting polymers. *Macromol. Rapid Commun.* **37**, 1623–1628 (2016).
- Langley, N. R. & Polmante, K. Relation of elastic modulus to crosslink and entanglement concentrations in rubber networks. *J. Polym. Sci. B* **12**, 1023–1034 (1974).
- Herbst, F., Döhler, D., Michael, P. & Binder, W. H. Self-healing polymers via supramolecular forces. *Macromol. Rapid Commun.* **34**, 203–220 (2013).
- Song, E. *et al.* Stretchable and transparent organic semiconducting thin film with conjugated polymer nanowires embedded in an elastomeric matrix. *Adv. Electron. Mater.* **2**, 1500250 (2016).
- Wu, H.-C. *et al.* A rapid and facile soft contact lamination method: evaluation of polymer semiconductors for stretchable transistors. *Chem. Mater.* **26**, 4544–4551 (2014).
- Chortos, A. *et al.* Highly stretchable transistor using microcracked organic semiconductor. *Adv. Mater.* **26**, 4253–4259 (2014).
- Kaltenbrunner, M. *et al.* An ultra-lightweight design for imperceptible plastic electronics. *Nature* **499**, 458–463 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by Samsung Electronics and the Air Force Office of Scientific Research (grant number FA9550-15-1-0106). S.R.-G. acknowledges the Fonds de Recherche Québecois, Nature et Technologie (FRQNT) for a postdoctoral fellowship. Y.-C.C. acknowledges the Ministry of Science and Technology, Taiwan, for partial financial support (project 104-2923-E-002-MY3). F.L. thanks the Swiss National Science Foundation for an Early Mobility Postdoc grant. B.C.S. acknowledges the National Research Fund of Luxembourg for financial support (project 6932623). J.L. acknowledges support from the National Science Foundation Graduate Research Fellowship Program under grant DGE-114747. T. Kurosawa acknowledges support from the Office of Naval Research (N00014-14-1-0142). X.G. acknowledges support from the Bridging Research Interactions through the collaborative Development Grants in Energy (BRIDGE) programme under the SunShot initiative of the Department of Energy (contract DE-FOA-0000654-1588). Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract DE-AC02-76SF00515. X-ray diffraction studies were performed at the Stanford Nano Shared Facilities.

Author Contributions J.Y.O., S.R.-G., Y.-C.C. and Z.B. conceived and designed the experiments. S.R.-G. and Z.B. designed the monomers and polymers. S.R.-G., B.C.S., T. Kurosawa, T. Katsumata and F.L. synthesized and characterized the monomers and polymers. J.Y.O. designed the device experiments and evaluated the stretchability of materials and devices. Y.-C.C. and J.Y.O. fabricated and optimized the OTFTs on solid substrates. J.Y.O. fabricated the fully stretchable OTFTs. J.Y.O., A.C., and C.Z. optimized the fully stretchable devices. J.Y.O. designed and performed the healing experiments. J.Y.O., G.-J.N.W., J.X., L.J., J.L., J.W.C., J.L. and Y.K. analysed the optical and mechanical properties of the polymer film. Y.-C.C. performed the grazing incidence X-ray diffraction experiments and analysis. Y.-C.C., X.G., S.R.-G., J.Y.O. and Z.B. proposed the new mechanism concept. J.Y.O. and W.-G.B. designed and drew the three-dimensional computer graphics. J.Y.O., S.R.-G., Y.-C.C., J.B.-H.T. and Z.B. co-wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.B. (zbao@stanford.edu).

METHODS

All details of materials synthesis and structural characterizations can be found in the Supplementary Information.

Materials synthesis and characterization. Synthesis of DPP-based conjugated polymers containing non-conjugated PDCA units in the backbone was carried out via Stille polymerization using different ratios of PDCA-containing monomers (see Supplementary Information). Branched alkyl chains were installed on the DPP monomer in order to increase solubility and π - π stacking. We also selected (E)-2-(2-(thiophenyl-2-yl)vinyl)thiophene (TVT) as co-monomer. The non-conjugated monomer was reacted with the brominated DPP monomer and the distannylated TVT under Stille polymerization conditions to afford **P1** to **P4**. Polymers were purified by precipitation in methanol followed by Soxhlet extraction using methanol, acetone, hexanes and chloroform. All polymers showed good solubility in different solvents such as chloroform and chlorobenzene. Increasing the non-conjugated moiety content within the backbone also improved the solubility in polar solvents such as THF and DMF. The synthetic procedures for precursors, monomers and polymers are detailed in the Supplementary Information. Molecular weight, polydispersity index, monomers and polymer characterization are all detailed in the Supplementary Information. All results are from polymers with number average molecular weight M_n over 14 kDa and with a polydispersity below 4. Polymers with low M_n values showed less improvement in stretchability.

Thin-film characterization. To be consistent with films used for transistor fabrication and characterization, all films tested have thicknesses in the range 30–35 nm. They were spin-coated on OTS-treated SiO_2/Si substrate from chlorobenzene solution (5 mg ml^{-1}) at 2,000 r.p.m. in glove box and thermally annealed at 150°C for 10 min. They were subsequently transferred to PDMS elastic substrates. The thin-film annealing temperature was found to affect stretchability. Significantly less improvement was observed for samples annealed at 150°C for 30 min. Therefore, annealing for 10 min was used for all films. The morphology of polymer films was measured by optical microscopy and AFM. The molecular stretchability of polymer film was investigated by polarized ultraviolet–visible spectroscopy. The elastic

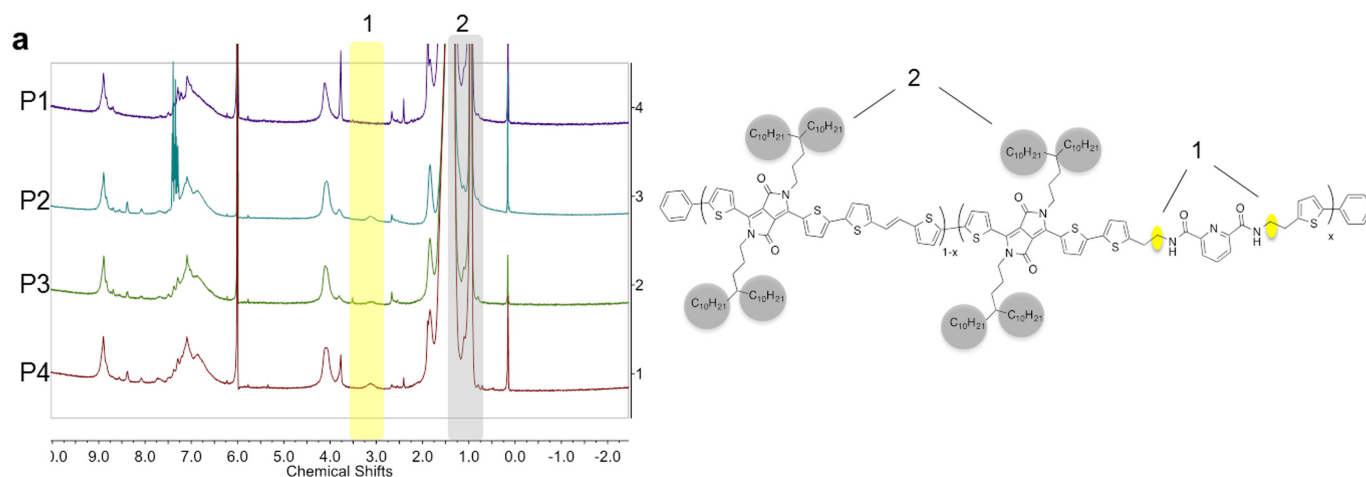
modulus of thin-films was measured using the buckling method (Supplementary Fig. 24). Grazing incidence X-ray diffraction was conducted at beamline 11-3 of the Stanford Synchrotron Radiation Lightsource. The incidence angle was 0.12° and the X-ray wavelength was 0.9758 \AA , corresponding to a beam energy of 12.7 keV . All grazing incidence X-ray diffraction images were collected in reflection mode with a two-dimensional area detector and the sample under a helium atmosphere. Stretching and relaxing rates of the polymer thin films were performed at 0.65 mm s^{-1} (strain rate $2.6\% \text{ s}^{-1}$).

Conventional organic transistors. The semiconducting polymer layer (35 nm) was spin-coated on OTS-treated SiO_2/Si substrate in glove box and was thermally annealed at 150°C for 10 min. The source and drain electrodes (Au, 40 nm) were thermally evaporated under $5.0 \times 10^{-6} \text{ Torr}$.

Skin-inspired organic transistor. Carbon nanotube (100 nm) film as the gate electronic was transferred onto polystyrene-block-poly(ethylene-ran-butylene)-block-polystyrene (SEBS) elastomer. The PDMS layer ($1.8 \mu\text{m}$) was then transferred onto a carbon nanotube/SEBS substrate. Then the semiconducting polymer layer (35 nm) was transferred on the PDMS dielectric layer. Finally, the PEDOT:PSS (30 nm)/carbon nanotube (70 nm) bilayer as the source and drain electrode was spray-coated onto the semiconducting layer. The transferring method for the device fabrication is depicted in Extended Data Fig. 5.

Healing process. For the solvent vapour treatment, the damaged sample is exposed to CHCl_3 vapour in a closed jar for 10 min at 40°C . For the heat treatment, the sample is thermally annealed on a 150°C hotplate for 30 min. For the combined solvent vapour and heat treatment, the sample first undergoes CHCl_3 vapour annealing (10 min at 40°C) and then thermal annealing (30 min on hotplate at 150°C).

31. Cellman, S. H., Dado, G. P., Liang, C.-B. & Adam, B. R. Conformation-directing effects of a single intramolecular amide–amide hydrogen bond: variable-temperature NMR and IR studies on a homologous diamide series. *J. Am. Chem. Soc.* **113**, 1164–1173 (1991).
32. Martin, R. B. Comparisons of indefinite self-association models. *Chem. Rev.* **96**, 3043–3064 (1996).

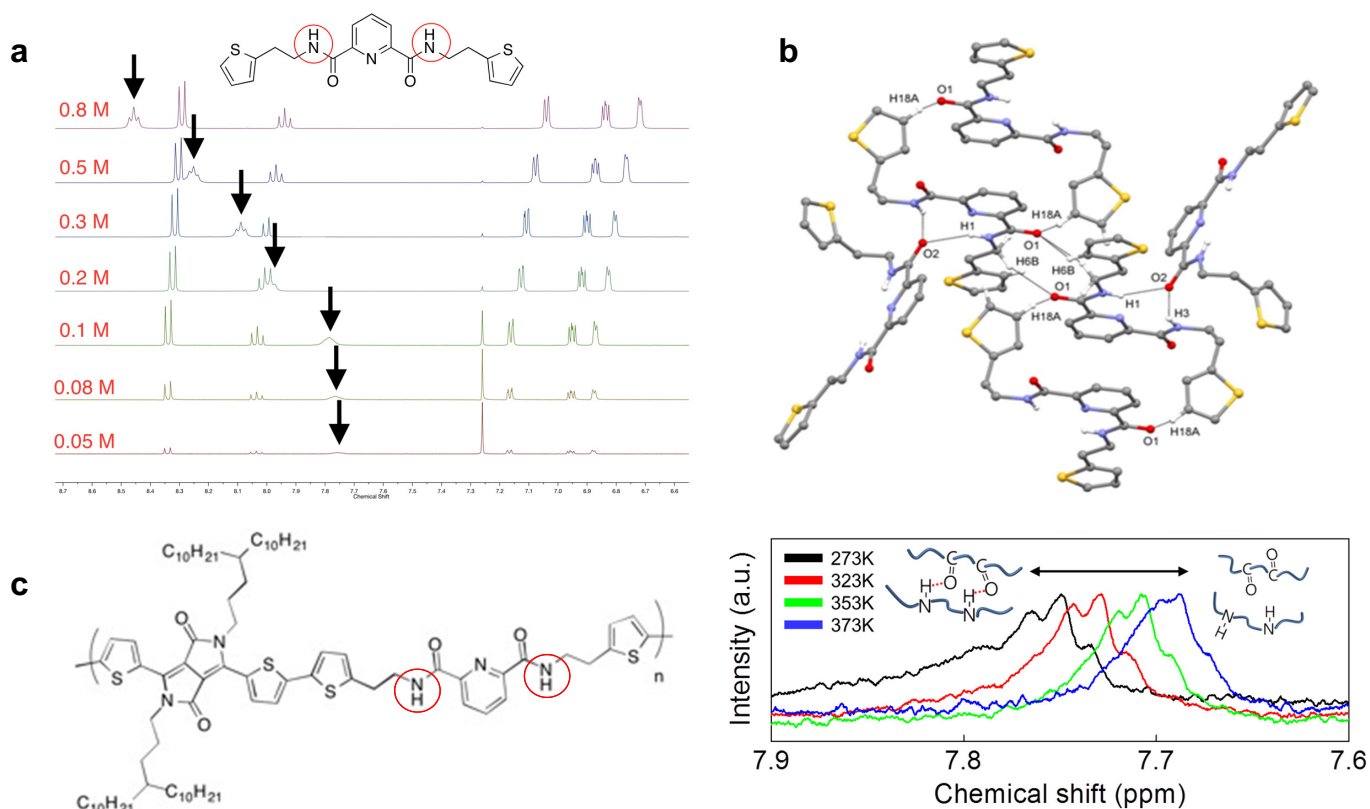


b

	Theoretical ratio (1:2)	Calculated ratio (1:2)	M_n (kDa)	M_w (kDa)	PDI	T_d^a	$\lambda_{\text{max}}^{\text{Sol}}$ (nm)	$\lambda_{\text{max}}^{\text{Film}}$ (nm)	HOMO ^b (eV)	Bandgap ^c (eV)
P1	0.00	0.00	20.4	65.4	3.2	>350	803, 738	801, 740	5.17	1.35
P2	0.05	0.06	19.7	61.4	3.3	>350	803, 738	801, 739	5.22	1.35
P3	0.10	0.09	16.2	49.1	3.0	>350	803, 738	801, 738	5.23	1.35
P4	0.20	0.17	13.4	35.5	2.6	>350	803, 738	801, 734	5.17	1.37

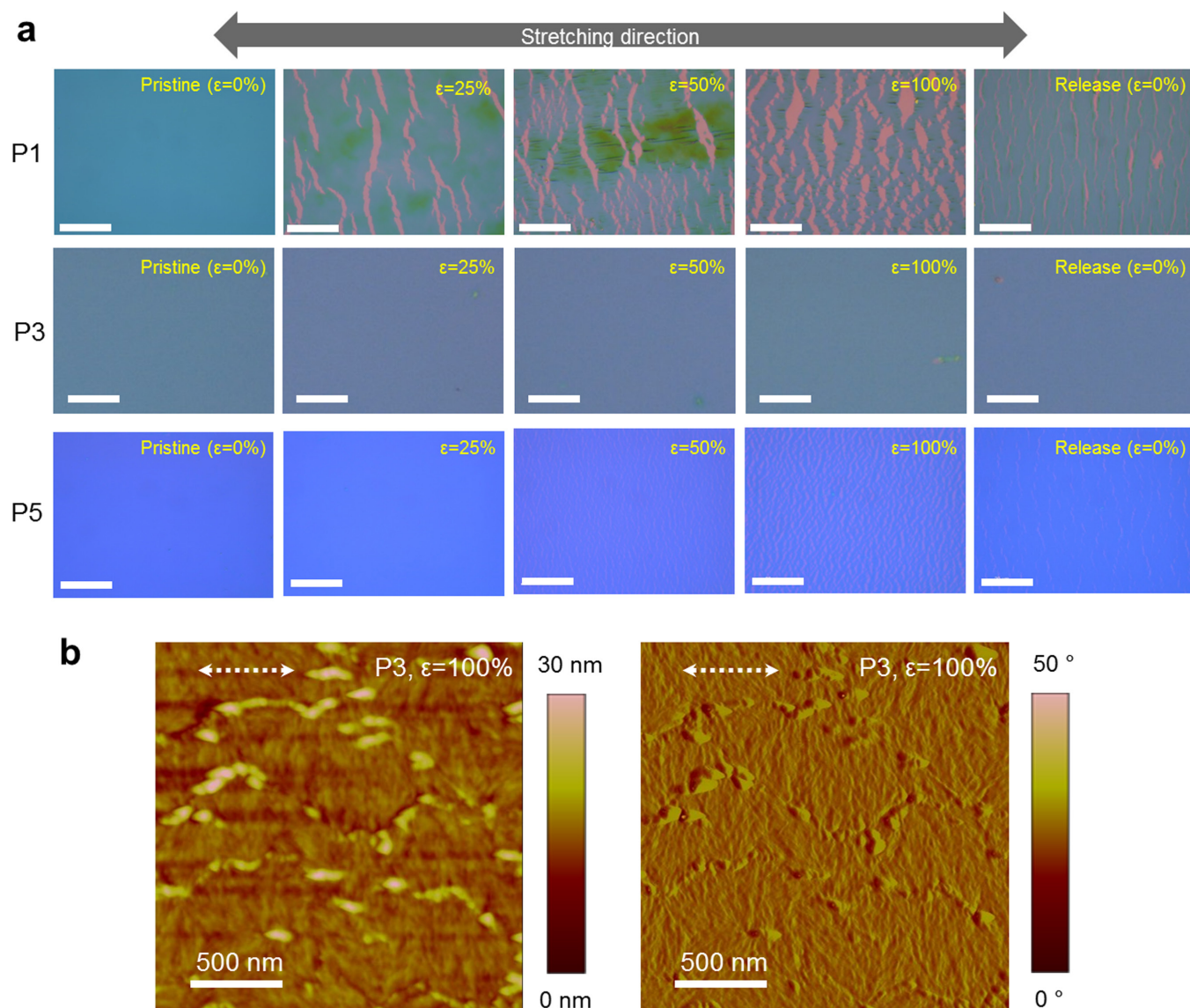
Extended Data Figure 1 | General characterization of P1–P4. **a**, Polymer composition by ^1H NMR. The ratio of PDCA moieties incorporated in the polymer backbone is determined by the integration of protons (1) versus the alkyl-chain terminal protons (2). **b**, General characterization of P1–P4. ^aDetermined from thermogravimetric analysis. ^bThe HOMO (highest occupied molecular orbital) energy level was calculated from

cyclic voltammetry. Potentials versus Ag/AgCl using 0.1 M TBAPF₆ (tetrabutylammonium hexafluorophosphate) in CH₃CN (acetonitrile) as the electrolyte solution. ^cCalculated using the following equation: gap = 1,240/ λ_{onset} of polymer film. (PDI, polydispersity; T_d , degradation temperature; M_w , weight average molecular weight; λ , absorption wavelength in ultraviolet/visible spectroscopy).



Extended Data Figure 2 | Intermolecular interactions between PDCA moieties. **a**, Chemical structure of model compound **1** and ^1H NMR at various concentrations of compound **1** in CDCl_3 . Upon increasing concentration from 0.05 M to 0.8 M, a distinct shift of the amide proton (black arrow) towards low fields is observed. This indicates hydrogen bonding between the PDCA moieties, as previously observed³¹. A dimerization constant of 0.18 M^{-1} was determined by plotting concentration versus chemical shift and fitting using a dimer association model³². **b**, Molecular structure of **1** showing intermolecular hydrogen bonds determined by single-crystal X-ray diffraction. Ellipsoids are set

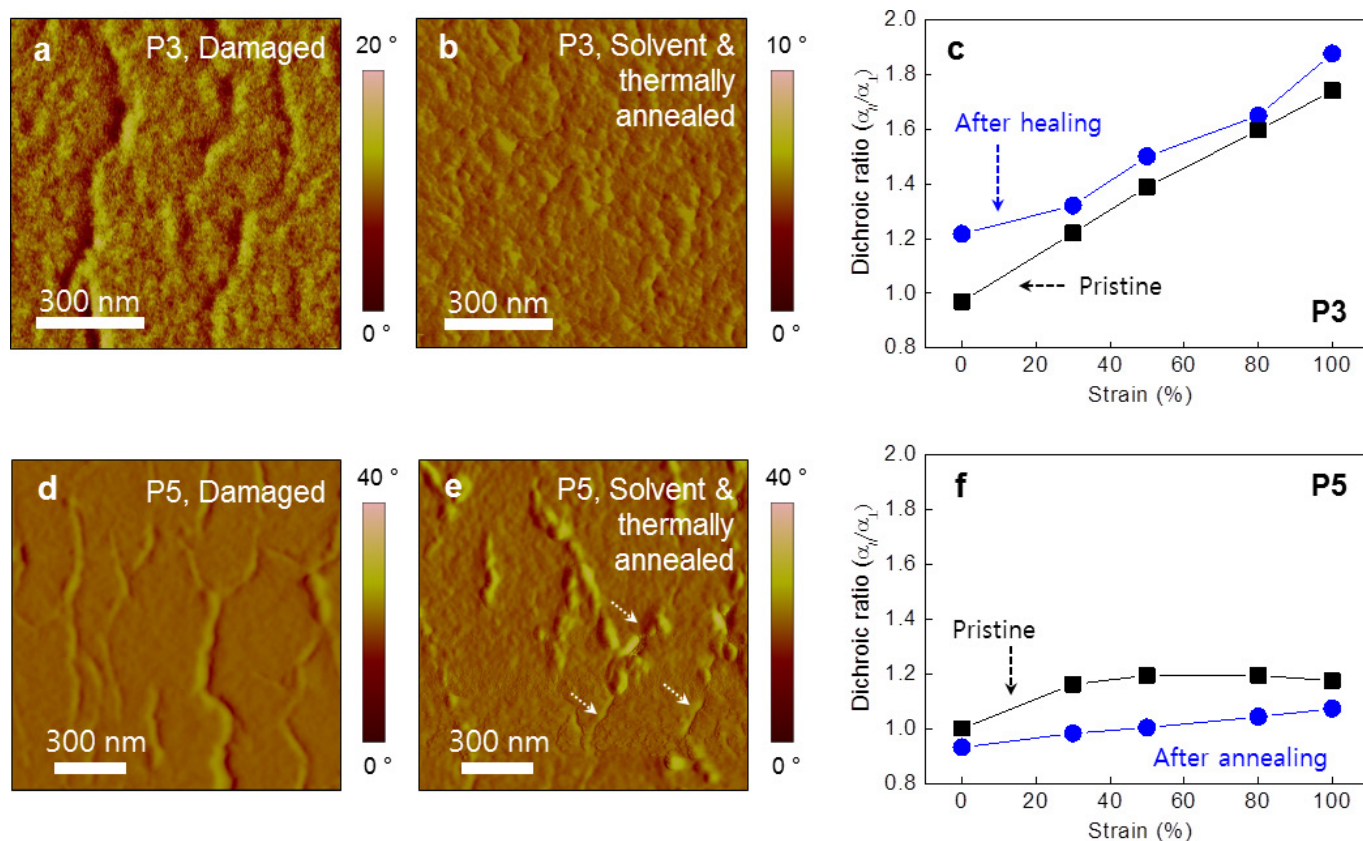
at the 30% probability level. Selected hydrogen atoms are omitted for clarity. **c**, Chemical structure of model oligomer **M1** and ^1H NMR of **M1** at various temperatures in $1,1,2,2\text{-tetrachloroethane-}d_2$ showing the amide NH protons peak. The chemical shift upon temperature increase indicates a breaking of the hydrogen bonds formed between the polymer chains (shown as insets). Oligomer **M1** was used for this study because the solubility of the polymer was not sufficiently high to perform a similar study under high concentrations. (a.u., arbitrary units; ppm, parts per million.)

**c**

Polymer	Applied Strain	FWHM at (200) peak <parallel>	Mean size of crystallites (nm)	FWHM at (200) peak <perpendicular>	Mean size of crystallites (nm)
P1	$\epsilon = 0\%$	0.0422	13.8	0.0417	14.0
	$\epsilon = 50\%$	0.0461	12.7	0.0489	11.9
	$\epsilon = 100\%$	0.0517	11.3	0.0554	10.5
	Release	0.0502	11.6	0.0525	11.1
P3	$\epsilon = 0\%$	0.0497	11.8	0.0494	11.8
	$\epsilon = 50\%$	0.0527	11.1	0.0526	11.1
	$\epsilon = 100\%$	0.0503	11.6	0.0524	11.2
	Release	0.0493	11.8	0.0491	11.9

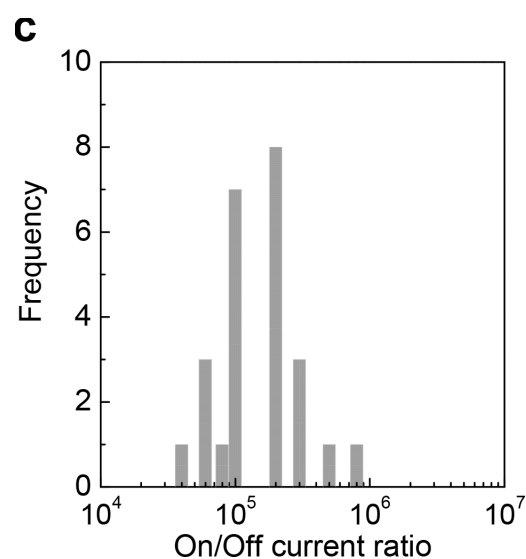
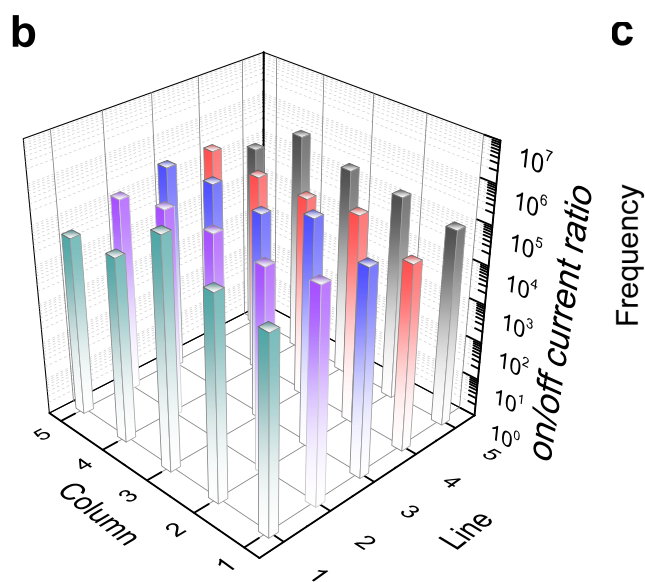
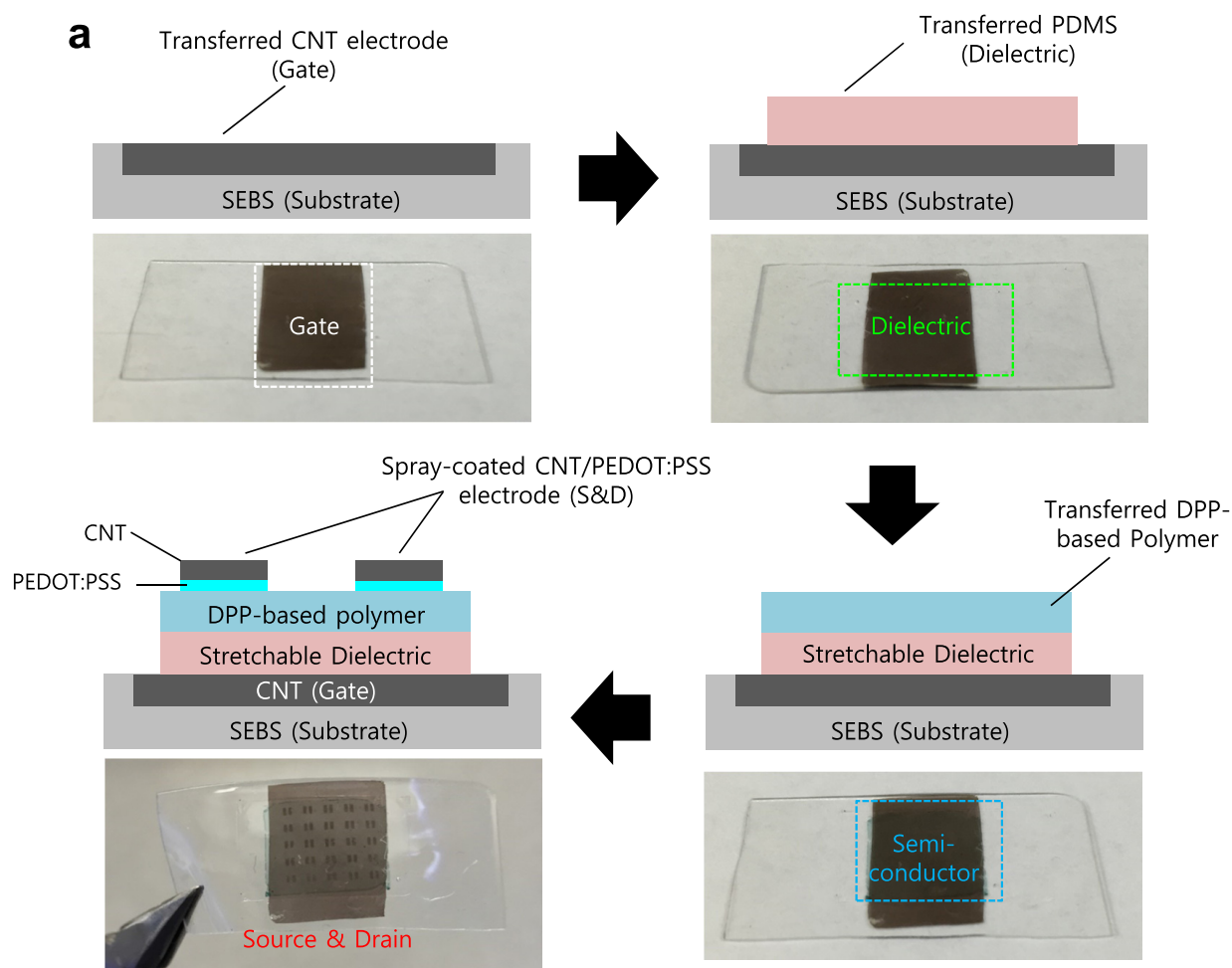
Extended Data Figure 3 | Behaviour of polymer thin-film under strain. **a**, Optical microscope images of **P1**, **P3** and **P5** as function of applied strain (0%–100%). **b**, Height and phase AFM images of **P3** under 100% strain showing no crack formation. **c**, Table of grazing incidence X-ray

diffraction data for **P1** and **P3** films as a function of strain (0%–100%). The samples are annealed at 150 °C for 10 min. A reduction in the mean size of crystallites is observed for **P1** to **P3**. The semiconductor is 35 nm thick. (FWHM, full-width at half-maximum.)



Extended Data Figure 4 | Thin-film characterization of damaged and healed P3 and P5. **a, b**, Atomic force microscopy images of damaged (**a**) and healed (**b**) thin film of **P3** after solvent and thermal annealing. We note that all the previously observed nanocracks were absent after the healing process. **c**, Dichroic ratio of **P3** healed thin film as determined by polarized ultraviolet–visible spectroscopy. **d, e**, Atomic force microscopy images of damaged (**d**) and healed (**e**) thin film of **P5** after solvent and thermal annealing. We observed that a small number of nanocracks

remained in the film. **f**, Dichroic ratio of **P5** healed thin film as determined by polarized ultraviolet–visible spectroscopy. We observed that the dichroic ratio of healed film of **P3** fully recovered to a value similar to that of the pristine film without damage. On the other hand, when **P5** is subjected to the same treatment, the dichroic ratio was not restored, indicating that the movement of the polymer chains was insufficient to restore the film's mechanical properties.



Extended Data Figure 5 | Fabrication and electronic properties of a fully stretchable 5×5 transistor array. **a**, Fabrication process of fully stretchable OTFTs. (1) Transfer printing of the carbon nanotube gate electrode as prepared by spray-coating a carbon nanotube solution (10 mg ml^{-1} in CHCl_3) on SEBS substrate (thickness $200 \mu\text{m}$). (2) Contact transfer printing of PDMS dielectric layer as prepared by spin-coating a diluted PDMS (220 mg ml^{-1} in CHCl_3) on OTS-treated SiO_2 substrate

on a carbon nanotube gate electrode. (3) Contact transfer printing of semiconducting polymer layer (prepared by spin coating onto OTS-treated SiO_2 substrate) on PDMS dielectric layer. (4) Spray coating of carbon nanotube (70 nm)/PEDOT:PSS (30 nm) source and drain electrodes on the semiconducting layer. **b**, **c**, On/off current mapping (**b**) and statistical distribution (**c**) of 5×5 fully stretchable OTFT arrays (width $1,000 \mu\text{m}$; length $150 \mu\text{m}$). Number of devices from a single substrate, $n = 25$.

Kalman Filter for Stripe Non-uniformity Correction in Infrared Focal Plane Arrays

ShuPeng Wang¹, ShiRu Zhang², NiZhuang Liu³

¹College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an, 710054, China

¹wang.shupeng@gmail.com

²zsr0504@qq.com

³College of Electircal and Control Engineering, Xi'an University of Science and Technology, Xi'an, 710054, China

³liuningzhuang@126.com

Abstract— This paper introduces a novel approach to correct stripe non-uniformity in infrared images. The differences between the bias of the column readout circuit of infrared focal plane arrays result in strong stripe noise. We propose that the bias of the column readout circuits can be observed by using a 1D bilateral filter on the average signal of the columns of infrared images. In this paper we regard the bias of column readout circuit as Gauss-Markov process. The proposed Gauss-Markov model provides a mechanism to capture the slow drift of the bias. A Kalman filter is derived that uses the observed bias to update the bias estimates. Finally, the infrared images are corrected using the estimated biases. The effectiveness of the method is shown according to the experimental results on real infrared images.

Keywords— Infrared image; stripe non-uniformity; Fixed pattern noise; Infrared focal plane array; Kalman filter

I. INTRODUCTION

Infrared imaging is very useful in industry, military and medical application since it can acquire infrared information. Infrared focal plane arrays (IRFPA) are employed predominantly in infrared imaging systems due to its numerous advantages such as compactness, production cost effectiveness, and high sensitivity[1]. An IRFPA is composed of small detectors and readout circuits. Because of mismatching of components and variations of parameters in the fabrication process, the spatial responses of pixel detectors and the bias voltages of readout circuits are different[2, 3]. Non-uniformity brings about the notorious non-uniformity noise, termed fixed pattern noise, which degrades quality of the acquired infrared images strongly. Non-uniformity correction (NUC) is a necessary step to improve infrared images[4]. Moreover, stripe non-uniformity which is due to non-uniformity of readout circuits is a special kind of non-uniformity, which should be considered as a special case[5, 6]. Since detector parameters drift slowly in time, one-time calibration of infrared focal plane array (IRFPA) cannot provide an effective solution to the NU problems. NUC must be performed repeatedly as drift occurs.

Non-uniformity correction techniques are categorized mainly into two classes: calibration-based and scene-based techniques. Calibration-based methods[7, 8] requires a uniform reference scene (typically, a blackbody radiation

source) in order to calibrate the response of each detector across the array so that all of detectors produce accurate and uniform readout values at the two different reference temperatures. This kind of NUC techniques often achieves radiometrically accurate readout values. However, since they need to interrupt the operation of the camera during the period when calibration is conducted, they are not convenient for real time applications. The calibration is applied in a long interval of time during which the parameters is considered as constant. Scene-based techniques rely on signal processing to remove the NU noise. They typically use an image sequence captured during the normal operation of the camera, and their performance relies on global motion in the sequence to provide diversity in the scene temperature per detector.

To date, numerous scene-based NUC techniques have been reported in the literatures. Harris and Chiang[9] developed algorithms that associate the mean and standard deviation of each detector with the gain and bias parameters by using the constant statistics assumption. This assumption postulates that the input irradiance is a uniformly distributed random variable and all pixel detectors should have the same mean and standard deviation. Under this assumption, the bias and gain of each detector can be estimated by employing linear mode. But this assumption is reasonable only if image sequence is long and so enough motion that each detector sweeps many different parts of the scene. Hayat et al.[10] proposed a finite impulse response (FIR) least-mean-square (LMS) filter that is used to estimate the gain and bias of each detector. The algorithm updates the gain and bias by use of subsequent images online and uses the updated parameters to correct infrared images. Torres et al.[11] developed a statistical technique called Constant Range that relies on the key assumption that all detectors are exposed to approximately the same range of irradiance. Ratliff et al. [12] developed an algebraic scene-based NUC technique that does not rely on any statistical or scene-diversity assumptions about the scene temperature. This algorithm employs estimates of interframe sub-pixel shifts in an image sequence and linear interpolation model for the motion to extract information on the bias non-uniformity. Torres and Hayat [13] model the gain and the bias of each detector as a Gauss-Markov random variable, and use a Kalman filter to update these estimates of the gain and bias

of each detector. Pezoa et al.[14] utilizes a bank of Kalman filters in parallel to estimate the gain and bias of each detector according to its own dynamic-model parameters. The final estimates of the gain and bias is achieved by forming a weighted superposition of all the estimates rendered by each Kalman filter.

Stripe noise is one of main fixed pattern noise caused by variation of the biases of column integration readout circuits in infrared imaging system[15, 16]. Since the detectors in the same column of IRFPA share the one column readout circuit and the detectors in the different column of IRFPA use different column readout circuit, the difference between the biases of the column readout circuits produces the different readout values at different columns when the detectors is exposed to the same radiation. The usual NUC techniques are not applicable to eliminate stripe noise. Stripe noise varies frame by frame, but calibration-based NUC methods cannot update parameters in time, which regard the parameters as invariable in a long time between two successive calibrations. The scene-based NUC techniques can resolve the problem that parameters drift slowly. However two problems may arise[17]: 1) a long image sequence is required for convergence. 2) a long image sequence may results in a “ghost artifacts” due to the previous images appears in the last one. In this paper, we propose a novel correction approach to eliminate stripe non-uniformity only using scene images. Assuming that the neighboring pixels in images are very similar, the difference between the pixels of neighboring column is caused mainly by biases of column readout circuit. Under this assumption, we proposed an approach to estimate the biases of readout circuit by using Kalman filter, and stripe non-uniformity is corrected using the estimated biases.

The paper is organized as follows. Section 2 presents the model for stripe non-uniformity of infrared images. The recursive estimate of the bias of column readout circuits is described in section 3. Experiments on real infrared image are described in section 4. Section 5 contains a conclusion.

II. MODEL FOR NU STRIPE NON-UNIFORMITY

Stripe non-uniformity in infrared images is due to difference between column integration readout circuits. We can make the following assumption:

- 1) The biases of column integration readout circuits can be regarded as unknown random variables. The pixels in the same column of IRFPA share the same bias.
- 2) The biases of different columns of IRFPA are independent of each other.

As a consequence of assumptions, the output of the $ijth$ detector of the kth frame is represented by

$$Y_k(i, j) = U_k(i, j) + X_k(j) + V_k(i, j), \quad (1)$$

$$i = 1, \dots, M, j = 1, \dots, N$$

where $X_k(j)$ is the bias of the jth column readout circuit in IRFPA. $U_k(i, j)$ is the infrared radiations collected by the

$ijth$ detector in the kth frame. $V(i, j)$ is the additive electronic noise. The sum of the value of infrared image $Y_k(i, j)$ along with column can be written as

$$Y_k(j) = U_k(j) + X_k(j) + V_k(j) \quad (2)$$

Assuming that the infrared radiations of neighbouring pixels in space are similar, the differences between columns of the image $Y_k(i, j)$ are mainly caused by the bias $X_k(j)$.

According to above assumption, $U_k(j)$ should be a slow changing signal, and $X_k(j)$ is considered as the additive temporal noise associated with the readout circuit of jth in infrared image.

We can separate the column signal $Y_k(j)$ into smooth part $Y_k^B(j)$ and detail part $Y_k^D(j)$ by applying a 1D low pass filter. The filtered output $Y_k^B(j)$ is computed as a weighted average of the signal $Y_k(j)$. The detail signal $Y_k^D(j)$ is obtained by subtracted $Y_k^B(j)$ from $Y_k(j)$ as following

$$Y_k^D(j) = Y_k(j) - Y_k^B(j) \quad (3)$$

The linear low-pass filter assigns weight coefficient only depending on spatial distance, which suffers with serious halo artifacts when there is a rapid variation of the signal. To solve this problem, our method adopts a nonlinear bilateral filter. The bilateral filter[18] are described by the following equation,

$$g(x) = \frac{1}{k(x)} \sum_{x'} f(x - x') h_s(\|x - x'\|) h_r(\|f(x) - f(x')\|) \quad (4)$$

where $k(x)$ is normalized factor, $h_s(x)$ is spatial kernel function and $h_r(x)$ is range kernel function. The key idea of bilateral filter is to avoid excessive smoothing of signal with rapid variation by using range kernel function, since $h_r(x)$ will assign a smaller weight coefficient with larger range difference between $f(x)$ and $f(x')$. The detail part $Y_k^D(j)$ of the column signal $Y_k(j)$ can be regard as the observation values of the bias of column readout circuits.

III. RECURSIVE ESTIMATION OF THE BIAS

The observed bias $Y_k^D(j)$ of the column readout circuits includes noise and observation error, so it can be written as

$$Y_k^D(j) = X_k(j) + N_k(j) \quad (5)$$

where $N_k(j)$ is an additive temporal noise with Gauss distribution and variance σ_N^2 . Observation vector corresponding to kth frame is expressed as

$Y_k^D = [Y_k^D(1), Y_k^D(2), \dots, Y_k^D(N)]^T$, which is an array of length N of column signal. Ultimately, we hope to achieve the minimum mean square error (MMSE) estimation of the state vector (bias vector) X_k from the observation sequences $\{Y_i^D\}_{i=0}^k$ in a recursive way.

Assuming that the variation of the bias of column readout circuits is a Gauss-Markov process, the bias X_{k+1} in $(k+1)th$ frame can be represented as autoregressive model,

$$X_{k+1} = A_k X_k + W_k \quad (6)$$

where $A_k = \text{diag}(a_k(1), a_k(2), \dots, a_k(N))$ is called the state transition matrix and $0 \leq a_k(i) < 1, i = 1, \dots, N$. W_k is gauss noise vector with identical and independent distribution, its covariance is $\text{diag}(\sigma_w^2, \dots, \sigma_w^2)$. The drift of the bias cannot lead to the long-term change of the dynamic range of the bias of readout circuits, so the average drift should be kept as zero. Assuming that the parameters of state transition matrix is same, i.e. $a_k(1) = \dots = a_k(N) = a$, the mean of noise W_k should be

$$M_k = E[W_k] = X_0(1-a) \quad (7)$$

where X_0 represents initial state.

Given the observation vector Y_k , the MMSE estimator of state vector X_k is represented recursively as

$$\hat{X}_k = \hat{X}_k^- + K_k [Y_k - \hat{X}_k^-] \quad (8)$$

where \hat{X}_k^- is called the priori estimate, defined as $\hat{X}_k^- \triangleq E[X_k | Y_1, \dots, Y_{k-1}]$, and can be solved iteratively using

$$\hat{X}_k^- = A_{k-1} \hat{X}_{k-1} + M_{k-1} \quad (9)$$

where M_{k-1} is given at Eq.(7). The matrix K_k is called Kalman gain matrix, and can be computed by using

$$K_k = P_k^- [P_k^- + \Phi]^{-1} \quad (10)$$

where Φ is the covariance matrix of observation noise. P_k^- is called a prior error covariance and defined as

$P_k^- \triangleq E[(X_k - \hat{X}_k^-)^2]$, which can be computed iteratively by using

$$P_k^- = A_{k-1} P_{k-1} A_{k-1}^T + Q_{k-1} \quad (11)$$

The error covariance matrix $P_k \triangleq E[(X_k - \hat{X}_k)^2]$ can be updated by using

$$P_k = (1 - H_k) P_k^- \quad (12)$$

To execute the above iterations, the initial conditions for state vector and error covariance matrix is given as following,

$$\hat{X}_0 = Y_0^D \quad (13)$$

$$P_0 = \text{diag}(\sigma_N^2, \dots, \sigma_N^2) \quad (14)$$

Finally, the observed infrared image $Y_k(i, j)$ is corrected

using the estimate bias \hat{X}_k ,

$$\hat{Y}_k(i, j) = Y_k(i, j) - \hat{X}_k(j) \quad (15)$$

IV. EXPERIMENTAL RESULTS

In this section, we discuss the experimental results obtained with the proposed algorithm. The data used in our experiment is realistic infrared images from a public infrared image dataset[19]. The experiments are carried out in MATLAB R2009a on a personal computer equipped with Windows (2.8 GHz Intel Core i5 CPU, 4 GB memory). The proposed algorithm is compared with the state of the art stripe noise removal method based on MHE[17, 20].

As shown in Fig.1, stripe noise in infrared images is reduced obviously using both MHE and our method. Although the MHE method reduce stripe noise, residual stripe noise is still obvious in images. Our noise removal results are better visually than MHE.

V. CONCLUSIONS

In this paper, we have presented a new technique to correct the stripe non-uniformity in infrared images. We take into account the fact that the stripe noise is produced by the non-uniform biases of column readout circuits in IRFPA, a novel approach is introduced to recursively estimate and correct bias non-uniformity. The proposed techniques works on average signal of the columns of infrared image. The observation of the bias of each readout circuit is achieved by using a 1D bilateral filter. Through modelling the bias by Gauss-Markov process, the proposed method captures the drift of the bias in each column. To achieve the recursive estimation, we derived a Kalman filter to estimate the bias of each column. Finally, we correct each detectors of infrared image using the estimated bias. The proposed algorithm is compared to the state of the art strip non-uniformity correction algorithm based on MHE using realistic infrared image. The experimental results show that our proposed algorithm obtains better effect than MHE.

ACKNOWLEDGMENT

This work was supported in part by the Science and Technology Plan of Shaanxi Province, China (Program No. 2014K06-37; 2015GY023), the Science and Technology Plan Industrial Application Technology Research and Development Project of Xi'an (Program No.CXY1440 (4)), the Xi'an City Beilin District Application Technology Research and Development Project (Program No.GX1417), the PhD starting capital of Xi'an University of Science and Technology (Program No.2015QDJ015).

REFERENCES

- [1] Martin, C.S., S.N. Torres, and J.E. Pezoa, Statistical recursive filtering for offset nonuniformity estimation in infrared focal-plane-array sensors. *Infrared Physics & Technology*, 2008. 58(6): p. 564-571.
- [2] Vera, E., P. Meza, and S. Torres, Total variation approach for adaptive nonuniformity correction in focal-plane arrays. *Optics letters*, 2011. 36(2): p. 172-174.
- [3] Cao, Y. and C.-L. Tisse, Single-image-based solution for optics temperature-dependent nonuniformity correction in an uncooled long-wave infrared camera. *Optics letters*, 2014. 39(3): p. 646-648.
- [4] Harris, J.G. and Y.-M. Chiang. Minimizing the "Ghosting" artifact in scene-based nonuniformity correction. in *SPIE Conference on Infrared Imaging Systems: Design Analysis, Modeling, and Testing IX* 1998. SPIE.
- [5] Zhao, J., et al., Single image stripe nonuniformity correction with gradient-constrained optimization model for infrared focal plane arrays. *Optics Communications*, 2013. 296: p. 47-52.
- [6] Qian, W., Q. Chen, and G. Gu, Minimum mean square error method for stripe nonuniformity correction. *Chinese Optics Letters*, 2011. 9(5): p. 051003.
- [7] Perry, D. and E. Dereniak, Linear theory of nonuniformity correction in infrared staring sensors. *Optical Engineering*, 1993. 32(8): p. 1854-1859.
- [8] Zhou, H., et al., Solution for the nonuniformity correction of infrared focal plane arrays. *Optical Engineering*, 2005. 44(15): p. 2928-2932.
- [9] Harris, J.G. and Y.-M. Chiang. Nonuniformity correction using constant average statistics constraint: analog and digital implementations. in *Proceeding of SPIE*. 1997. SPIE.
- [10] Hayat, M.M., et al., Statistical algorithm for nonuniformity correction in focal plane arrays. *Applied Optics*, 1999. 38(8): p. 772-780.
- [11] Torres, S.N., R.A. Reeves, and M.M. Hayat. Scene-based nonuniformity correction method using constant-range: performance and analysis. in *Proceedings of 6th world multiconference on systemics, cybernetics and informatics*. 2002. Orlando, FL, USA.
- [12] Ratliff, B.M., M.M. Hayat, and R.C. Hardie, An algebraic algorithm for nonuniformity correction in focal plane arrays. *the Optical Society of America A*, 2002. 19(9): p. 1739-1747.
- [13] Torres, S.N. and M.M. Hayat, Kalman filtering for adaptive nonuniformity correction in infrared focal-plane arrays. *the Optical Society of America A*, 2003. 20(3): p. 470-480.
- [14] Pezoa, J.E., et al., Multimodel kalman filtering for adaptive nonuniformity correction in infrared sensors. *Journal of the Optical Society of America*, 2006. 23(6): p. 1282-1291.
- [15] Sui, X.-B., Q. Chen, and G.-H. Gu, Algorithm for eliminating strip noise in infrared image. *Journal of Infrared Millimeter waves*, 2012. 31(2): p. 106-112.
- [16] Qian, W., et al., Correction method for stripe nonuniformity. *Applied optics*, 2010. 49(10): p. 1764-1773.
- [17] Tendero, Y., et al. Efficient single image non-uniformity correction algorithm. in *Proceedings of SPIE*. 2010. Bellingham, WA.
- [18] Tomasi, C. and R. Manduchi. Bilateral filtering for gray and color images. in *Computer Vision*, 1998. Sixth International Conference on. 1998. IEEE.
- [19] Tendero, Y. and J. Gilles. ADMIRE: a locally adaptive single-image, non-uniformity correction and denoising algorithm: application to uncooled IR camera. in *SPIE Defense, Security, and Sensing*. 2012. International Society for Optics and Photonics.
- [20] Delon, J., Midway image equalization. *Journal of Mathematical Imaging and Vision*, 2004. 21: p. 119-134.

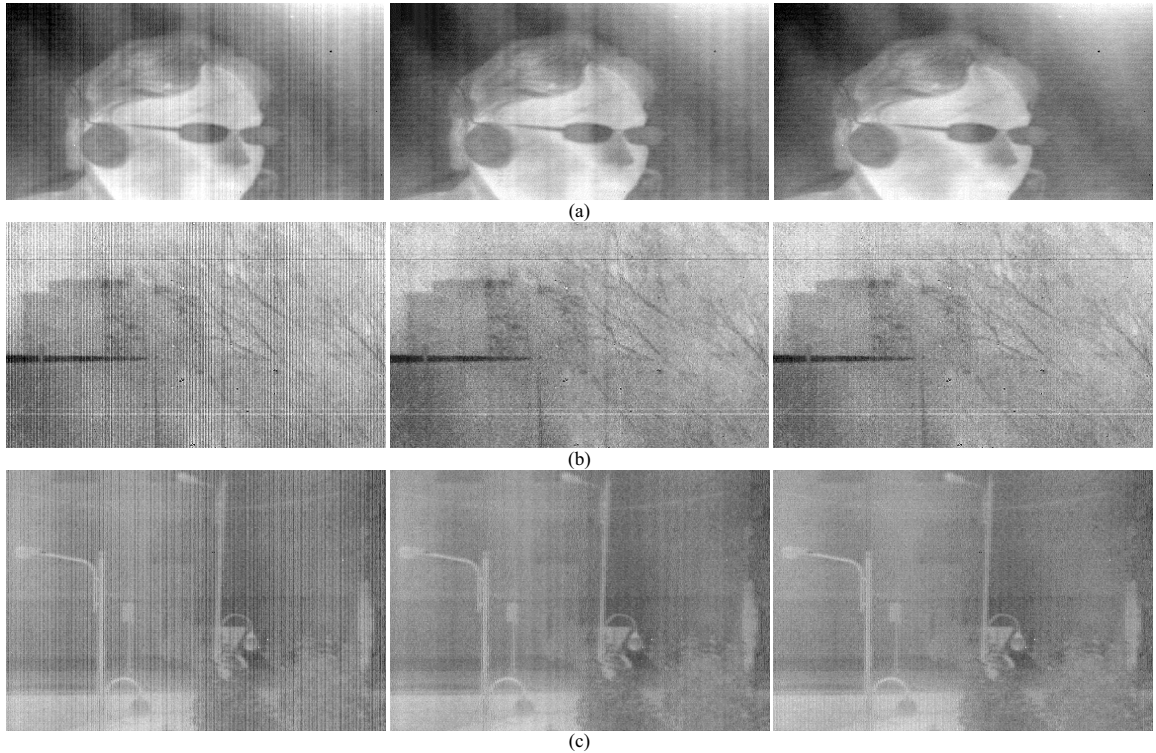


Fig. 1 Stripe noise removal results based on MHE (the second column images) and our method(the third column images)

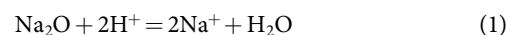
Implications for metal and volatile cycles from the pH of subduction zone fluids

Matthieu E. Galvez¹, James A. D. Connolly¹ & Craig E. Manning²

The chemistry of aqueous fluids controls the transport and exchange—the cycles—of metals^{1–5} and volatile elements^{3,6,7} on Earth. Subduction zones, where oceanic plates sink into the Earth's interior, are the most important geodynamic setting for this fluid-mediated chemical exchange^{2,6–10}. Characterizing the ionic speciation and pH of fluids equilibrated with rocks at subduction zone conditions has long been a major challenge in Earth science^{11,12}. Here we report thermodynamic predictions of fluid–rock equilibria that tie together models of the thermal structure, mineralogy and fluid speciation of subduction zones. We find that the pH of fluids in subducted crustal lithologies is confined to a mildly alkaline range, modulated by rock volatile and chlorine contents. Cold subduction typical of the Phanerozoic eon¹³ favours the preservation of oxidized carbon in subducting slabs. In contrast, the pH of mantle wedge fluids is very sensitive to minor variations in rock composition. These variations may be caused by intramantle differentiation, or by infiltration of fluids enriched in alkali components extracted from the subducted crust. The sensitivity of pH to soluble elements in low abundance in the host rocks, such as carbon, alkali metals and halogens, illustrates a feedback between the chemistry of the Earth's atmosphere–ocean system^{14,15} and the speciation of subduction zone fluids via the composition of the seawater-altered oceanic lithosphere. Our findings provide a perspective on the controlling reactions that have coupled metal and volatile cycles in subduction zones for more than 3 billion years⁷.

Water-based chemical reactions are essential to Earth's surface chemistry. Water is also crucial to our planet's inner chemical dynamics. Subduction zones are the primary locations where surface water is delivered to the deep Earth, promoting low-temperature (T) melting of the crust¹⁶ and mantle⁹, affecting the rheology of the lithosphere¹⁷, blurring pre-metamorphic signatures of ancient Earth's palaeo-environmental conditions^{10,18} and contributing to the global recycling of volatiles^{6–8} and metals^{2,6,8}. Consequently, characterizing the reactivity of supercritical aqueous solutions in subduction zones is a problem of both scientific and societal relevance in Earth science. The key variables that control rock mineralogy and fluid chemistry are pressure (P), T and composition, which in combination govern, for example, fluid redox state⁶, speciation⁵ and electrical properties¹¹. Although the redox budget of subduction zones has been the subject of growing attention⁶, obtaining quantitative constraints on their acid–base budget and on the ionic (for example, the pH; refs 5, 12) and molecular composition of fluids at mantle conditions has proven difficult in both theory and experiment^{11,12}. Recent molecular dynamic simulations have extended the range of conditions at which the solvent property of water (its dielectric constant ϵ_r , Extended Data Fig. 1) is known to upper-mantle conditions^{19,20}, and these results have been generalized for geologically ubiquitous carbonic fluids^{21,22}. However, geochemical applications remain hindered by insufficient knowledge about the sensitivity of fluid properties to changing mineralogical assemblages at the elevated P and T prevalent in the Earth's interior.

We address this problem using a thermodynamic approach²² that connects fluid composition and speciation to predicted rock mineralogy as a function of P and T . After determining the P – T stability fields of solid–solvent phase assemblages for crustal and mantle lithologies (Fig. 1, Methods and Extended Data Table 1), the chemical potentials from the mineralogical model are used to derive fluid speciation (Figs 2 and 3) using oxide hydrolysis equilibria represented, for example, by the alkali – proton exchange reaction:



We assume fluid–rock equilibrium holds at all conditions, as supported by rapid rates of mineral dissolution and precipitation in hydrothermal fluids²³; our results hold for common scenarios of pervasive fluid flow at low transport rates where grain boundary diffusion dominates. Redox reactions of Fe have been discounted, that is, the redox composition of the fluid (O/H ratio) is determined by the devolatilization process. At equilibrium, under the assumption of rock-dominated chemistry, the chemical potential (μ) of oxides constrains fluid speciation by:

$$\underbrace{\mu_{\text{Na}_2\text{O}} - \mu_{\text{H}_2\text{O}}}_{\text{mineralogical model}} = 2 \underbrace{(\mu_{\text{Na}^+} - \mu_{\text{H}^+})}_{\text{solute speciation}} \quad (2)$$

The partial molar Gibbs energy of the i th solute is $\mu_i = g_i^\ominus + RT \ln(\gamma_i m_i)$, where g_i^\ominus is the standard-state Gibbs energy of 1 mol of solute per kg of solvent at P and T . Here γ_i is the solute activity coefficient, m_i the molal concentration, their product $\gamma_i m_i$ is the activity (a) of i , and $\text{pH} = -\log_{10}(a_{\text{H}^+})$. Thermochemical data for minerals, solvents and solutes are reported in Methods and Extended Data Table 2. This approach provides fluid speciation and pH by charge balance (Methods), and is original in that it tracks the release of all soluble elements simultaneously. We focus on pH, and particular emphasis is placed on testing the sensitivity of predictions to change in P – T conditions and rock composition (Figs 2 and 3).

The fluid alkalinity is defined by $\Delta\text{pH} = \text{pH} - \text{pH}_n$, which tracks the acidic ($\Delta\text{pH} < 0$) or alkaline ($\Delta\text{pH} > 0$) character of the fluid with respect to acid–base neutrality. Neutral pH ($\text{pH}_n = 0.5[\text{p}K_w - \log_{10}(a_{\text{H}_2\text{O}})]$) is defined as the condition where $a_{\text{H}^+} = a_{\text{OH}^-}$, and, accordingly, model ΔpH takes into account the modification of water activity and ionization constant (K_w) in the mixed solvents. Crustal pelite and basalt exhibit similar isothermal trends of ΔpH when H_2O is the solvent. This pattern reflects uniform variations in partial molar volume of Na_2O in these systems where sodic phases include feldspar (albite, Ab), mica (paragonite, Par), clinopyroxene (jadeite, Jd) and amphibole (Fig. 1a, b). The ΔpH reaches a maximum at conditions corresponding to the $\text{Ab} = \text{Jd} + \text{Quartz (Q)}$ reaction, and is in the range 2.3 ± 0.2 along representative cold and intermediate P – T paths (Figs 1, 2d). The ΔpH of fluids equilibrated with an alkali-free mantle peridotite spans a larger range. It peaks near 2.1 ± 0.2 , reflecting incompatible behaviour of calcium in the ultramafic system, at 400 – 500°C ,

¹Earth Sciences Department, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland. ²Department of Earth, Planetary and Space Sciences, University of California, Los Angeles, California 90095-1567, USA.

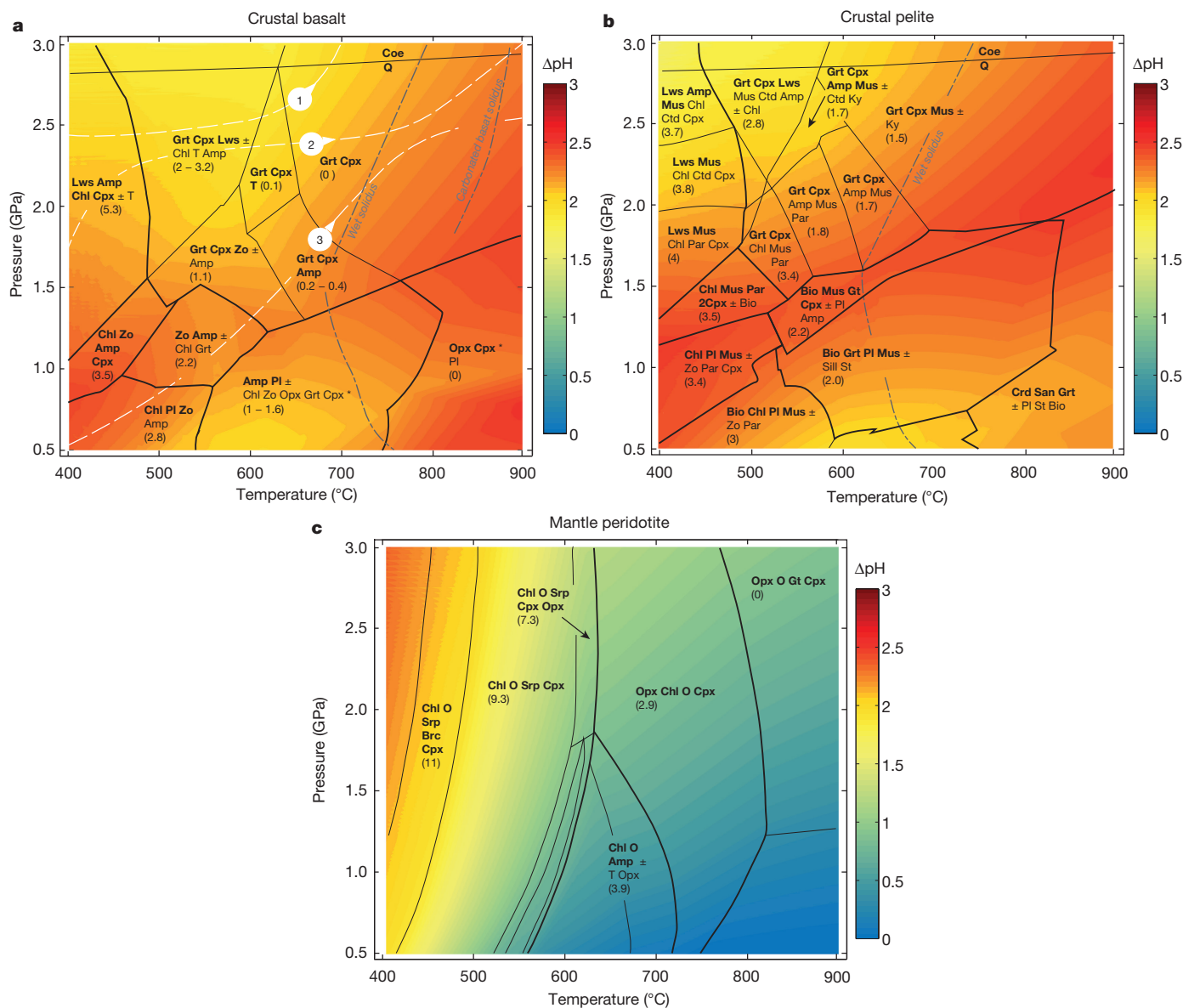


Figure 1 | $\Delta p\text{H}$ of solutions in equilibrium with basalt, pelite and peridotite lithologies. Here $\Delta p\text{H} = p\text{H} - p\text{H}_n$ describes the deviation of fluid pH from pH at acid–base neutrality, $p\text{H}_n$ (see text and Methods). It is plotted against the stable mineral assemblages (that label each field) as a function of P and T for basalt (a), pelite (b) and peridotite (alkali-free lherzolite) (c) compositions (Extended Data Table 1) at equilibrium with H_2O . Mineral abbreviations (defined in Methods) in bold denote phases present in the entire field, those in plain text appear in some parts of the field. Indicative amounts of H_2O structurally bound in minerals are given in parentheses in wt% below the characteristic phase assemblage. Decreasing amount of H_2O in the mineral assemblage with heating indicates progressive rock dehydration. The carbonated basalt (Fig. 2)

and declines to about 1.0 with heating and transition to chlorite- and garnet-peridotite along cold and intermediate P - T paths (Figs 1c, 2d). These trends are maintained along the path representative of hot Precambrian subduction zones¹³ (Figs 1a, 2a).

Carbonic solvents formed if graphite or carbonate are present in the pelite (redox neutral solvent with atomic ratio $\text{O}/\text{H} = 1/2$) and basalt (oxidized, $\text{O}/\text{H} > 1/2$), respectively, are characterized by lower H_2O activity ($0.6 \leq a_{\text{H}_2\text{O}} < 1$). The $\Delta p\text{H}$ of the C-bearing solutions decreases by up to 0.7 units for the carbonated-basalt fluids (Fig. 2b, e). Their carbon abundance (Fig. 3a) is about one order of magnitude greater than in fluids equilibrated with graphitic pelite¹⁰. Although metal solubilities are semiquantitative at these conditions^{21,24}, the $(\text{Na} + \text{K})/\text{Al}$

differs from the C-free system by the presence of calcite/aragonite (up to ~ 1.3 GPa), dolomite (~ 1 – 1.9 GPa) and magnesite above ~ 1.9 GPa. Two Phanerozoic P - T paths (white dashed lines), that is, Honshu ('cold' path, labelled 1) and Columbia ('intermediate' path, 2) are indicated in a along with a 'hot' P - T path (path 3) (Methods). Dash-dotted curves in grey are approximate experimental locations of the wet basalt/pelite solidi in a and b, and also show the carbonated-basalt solidus in a (Methods). All fields for basalt and pelite, with the exception of those marked with an asterisk where quartz is not stable at $P < 0.7$ GPa and $T > 600$ °C, contain quartz or coesite (Coe). Thick lines designate boundaries between classical metamorphic facies referred to in Fig. 2. Precision on absolute $\Delta p\text{H}$ values is ± 0.2 $\Delta p\text{H}$. Accuracy is within 1 $\Delta p\text{H}$ unit (Methods).

ratios of fluids increases, by a factor of 2 to 20 for pelite and basalt, respectively, with minor bulk-rock solubility increase. These modifications are due to competing solvent (H_2O dilution with non-dipolar molecules such as CO_2 and CH_4), pH (carbonate equilibria and equation (2)), and speciation effects (alkali-carbonate ion pairing). It illustrates that the ability of subduction zone fluids to dissolve metals, irrespective of their accessible oxidation state, depends on the fluid's volatile speciation, and thus redox state. Similarly, the solubility of C and the C/Ca ratio in the basaltic fluid increase with heating and decompression (Fig. 3a, b). These trends are supported by experimental evidence²⁵, and controlled by incongruent dissolution of Ca-Mg-Fe carbonate. Along hot to intermediate geotherms, crustal and oceanic

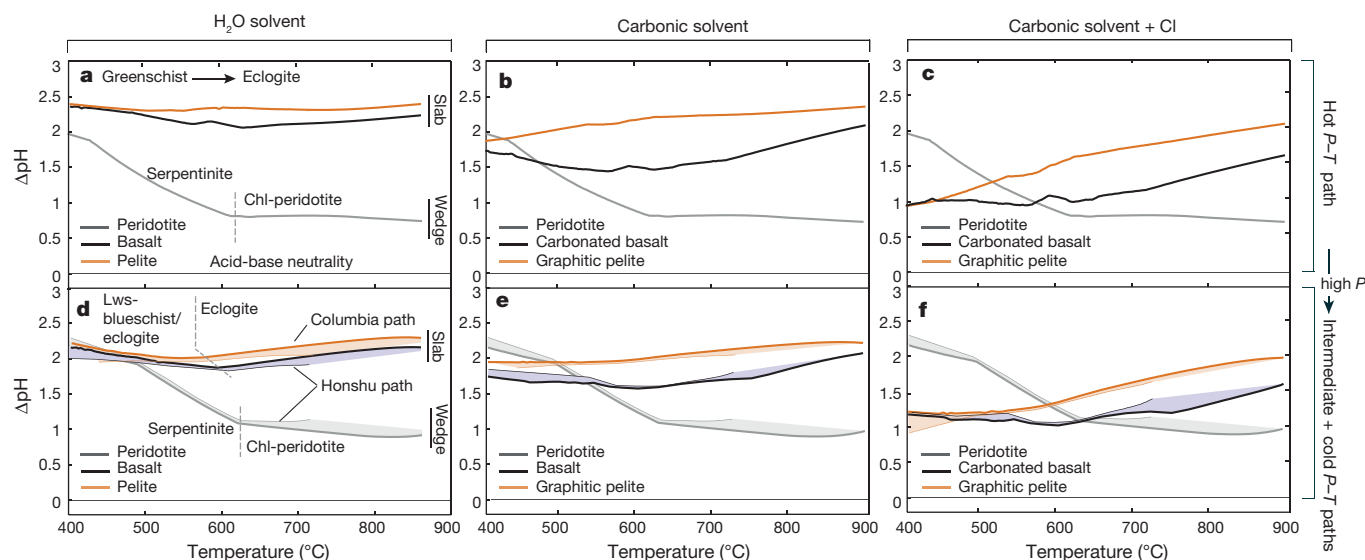


Figure 2 | ΔpH along three representative Precambrian and Phanerozoic P - T paths. The three paths are shown in Fig. 1a. **a–f**, ΔpH of a pure H_2O solvent (**a**, **d**) at equilibrium with peridotite (grey), basalt (red) and pelite (yellow) along hot (**a–c**), intermediate and cold P - T paths (**d–f**) (see Methods for details). **b**, **e**, Results obtained for a carbonic solvent, that is, with C added to the pelite (redox-neutral solvent, $a_{\text{C}} = 1$, $\text{O}/\text{H} = 0.5$ (atomic ratio) and $a_{\text{CO}_2} \approx a_{\text{CH}_4}$) and basalt (H_2O - CO_2 solvent, $\text{O}/\text{H} > 0.5$). Greater C solubility and CO_2 ionization (carbon species include, CaHCO_3^+ , CaCO_3^0 , HCO_3^- , NaCO_3^- , CO_3^{2-}

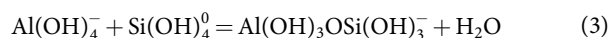
$\pm \text{HCOO}^-$, Extended Data Table 2) account for larger ΔpH drop in the carbonated basalt compared to the graphite-bearing pelite system. **c**, **f**, Results obtained for carbonic solvents including 1 molal total chlorine (Methods). The effect of chlorine is most pronounced at low- T conditions where it is fully ionized, and its abundance in the fluid exceeds the ionic strength of Cl-free solutions, $I \approx 0.1$, by ~ 1 order of magnitude. Major metamorphic facies transitions are indicated in **a**, **d**, and acid-base neutrality ($\Delta\text{pH} = 0$) is indicated by the horizontal baseline in each panel. The P - T paths refer to those defined in Fig. 1.

mantle dehydration occur at shallow depths (< 3 GPa ≈ 80 km, Fig. 1) and are concomitant with elevated C solubility (4–10 mol C per kg H_2O , Fig. 3b), favouring ample fore-arc carbonate dissolution and melting (Methods, Extended Data Table 3). This regime characterizes hot Precambrian subduction zones and rare modern equivalents. The cold path differs fundamentally in that slab dehydration is limited, and confined to elevated P (Fig. 1a) and low T where the solubility of carbon is relatively low (0.3–3 mol per kg H_2O ; Fig. 3a). Thus, cold paths typical of the Phanerozoic favour the mechanical incorporation of carbonates in the deep upper mantle (Methods, Extended Data Table 3), in agreement, for example, with the ubiquitous persistence of Ca-Mg-Fe carbonates in ultra-high- P terranes on Earth²⁶. This temporal directionality implies that the return of surface O to the mantle by slab carbonates is enhanced by subduction cooling over geological time (Extended Data Table 3). The formation of extensive domains of mantle redox heterogeneities, and their associated processes^{10,27}, may be a geologically recent phenomenon.

Addition of chlorine in amounts typical of dilute subduction zone fluids²⁸ shifts predicted ΔpH closer to neutrality regardless of thermal structure (Fig. 2c, f). Thus, at $T > 500$ °C, the ΔpH of the three studied lithologies is confined to a narrow mildly alkaline range (Fig. 2). Larger contrasts develop at lower temperatures, or at elevated T and P lower than 1 GPa (Fig. 1) characterized by enhanced ion association and complex neutralization, and, thus, pH is a less pertinent system variable. The ΔpH is particularly sensitive to seawater components in the altered oceanic lithosphere such as carbon, alkali and halogen (Fig. 2). Slab thermal structure plays a subordinate role in determining pH, but exerts an important control on the geochemical systematics of the volatile-rich fluids (Fig. 3a, b).

These findings are also illustrated by the evolution of fluid chemistry across a chemical profile ranging from an alkali-depleted mantle composition to a model slab fluid composition (component \mathcal{F} in Fig. 3c, d and Extended Data Table 1). This scenario tests the role of chemical variability, introduced by metasomatism or intramantle differentiation, on fluid chemistry. At 600 °C, 2 GPa and minor rock Na_2O content ($w_{\text{Na}_2\text{O}}^{\text{rock}}$) of 0–0.1 wt% (in clinopyroxene), that is, alkali

compositions typical of depleted supra-subduction-zone mantle peridotite²⁹, the pronounced increase of $\mu_{\text{Na}_2\text{O}}$ (Extended Data Fig. 2) is not fully accommodated by enhanced deprotonation of, for example, $\text{Si}(\text{OH})_4^0$ to $\text{SiO}(\text{OH})_3^-$ (Extended Data Fig. 3) and leads to a residual ΔpH increase exceeding 2 log units (Fig. 3c, d). The variations deviate significantly from a linear mixing trend between endmember fluid compositions, and highlight the limited acid-base buffering capacity of ubiquitous ultramafic mineral assemblages characterized by low alkali concentrations. Previous models of nitrogen speciation in the mantle³⁰ may be revised in light of these new constraints. The formation of Na,K-bearing assemblages including clinopyroxene, biotite and amphibole stabilizes ΔpH at strongly alkaline values (as previously anticipated by ref. 31)— $\Delta\text{pH} \geq 3$ –4 over the interval 400–900 °C at 2 GPa; pH values then gently decline along a sub-linear mixing trend with further addition of component \mathcal{F} . Low values of μ_{SiO_2} in mantle peridotites (Extended Data Fig. 2) would act as a geochemical trap for a range of silica-bearing polynuclear clusters that carry trace elements in metasomatic fluids derived from the slabs^{24,32} (Extended Data Fig. 3). The formation of mica (at $X_{\mathcal{F}} = 0.43$, Fig. 3d) causes a drop of ΔpH , and a coincident increase of μ_{SiO_2} , $\mu_{\text{Al}_2\text{O}_3}$ (Extended Data Fig. 2). These conditions would promote the high- T formation of silica–aluminate complexes³³, not considered here for lack of relevant thermochemical data, by condensation reactions such as:



We anticipate that the formation of these solutes would increase the solubility of cations (as required from electroneutrality) and bring ΔpH values closer to acid–base neutrality. Further work is needed to quantify the efficiency of this mechanism. Experimental strategies for pH and electrical transport¹¹ measurements in high- P fluids are required.

Taken together, our results show that acid–base mechanisms important for the dissolution, transport and deposition of metals, for example pH fluctuations, are not restricted to upper-crustal magmatic–hydrothermal environments^{4,5}—they are significant in

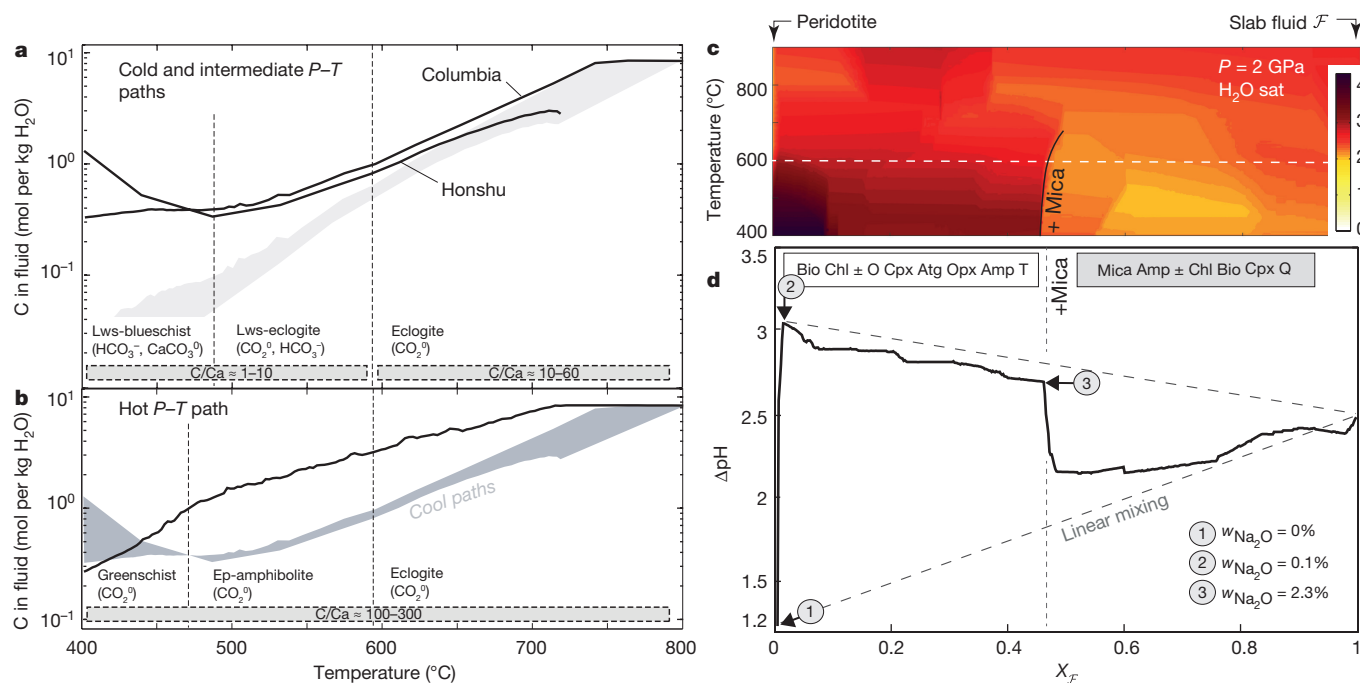


Figure 3 | Metasomatism at the subduction interface. **a**, C content of a fluid at equilibrium with carbonated basalt along cold and intermediate P - T paths. **b**, C content of the fluid along the hot P - T path. C abundance is displayed here in moles of C per kg of H_2O . Atomic ratio C/Ca along those paths is indicated. It is particularly sensitive to the P - T conditions, rising very rapidly to above 10^3 with heating and decompression. Enhancement of C solubility with pressure, a peculiar feature confined to $P > 2.5$ GPa and $T < 500^\circ\text{C}$ (The stable carbonate phase is Fe-Mg carbonate 'Magnesite'), is due to extensive Ca-carbonate complexing in the fluid. Predictions from simple molecular fluid models are indicated along the same paths (pale grey area, **a**). Discrepancies between the models are as expected largest at low T and elevated P . Vertical dashed lines indicate

boundaries of metamorphic facies (for example, lawsonite (Lws)-eclogite), and indications on C speciation (for example CO_2 , HCO_3^-). **c**, Variations of ΔpH (colour coded) for $T = 400$ – 900°C and $P = 2$ GPa obtained by mixing peridotite composition with model metasomatic component $\mathcal{F} = \text{Na}(\text{K}, \text{Ca}, \text{Mg}, \text{Fe})_{1/3}\text{Al}_{1.5}\text{Si}_2\text{O}_{17.9}$ (expressed in mole fraction of the ultramafic- \mathcal{F} oxide mixture, $X_{\mathcal{F}}$) at H_2O saturation (H_2O sat.). Thick black line indicates appearance of white mica (+Mica) in the assemblage. **d**, Section plot of **c** at 600°C with indications of mineral assemblages and corresponding values of Na_2O abundance in the rock (arrows, $w_{\text{Na}_2\text{O}}^{\text{rock}}$ in wt%). Dashed grey lines indicate linear mixing between endmember fluid compositions.

polar solvents present at the slab interface. Factors controlling these fluctuations include (1) mineralogical heterogeneities, (2) the presence of minor but soluble components in solid solutions, (3) variations of the activity and structure of H_2O in mixed-volatile solvents, and (4) the speciation of coordinating ligands. In particular, sulfur may form abundant dipolar H_2S and SO_2 molecules and various ligands, the distribution of which is redox and pH sensitive³ and will influence the transport of metals such as Au and Cu along the subduction interface. Our work places quantitative pH restrictions on the identity and solubility of aqueous metal-complexes pertinent to these processes. The development and generalization of acid-base formalisms to coexisting rock–fluid–melt systems will be necessary at T greater than 750 – 800°C (Fig. 1a, Methods).

The present work focused on fundamental chemical characteristics of aqueous fluids at upper-mantle conditions, which improves our ability to interpret compositional variability in rocks exhumed from the Earth's interior^{18,21}. Metal–volatile chemical coupling is a consequence of mineralogical equilibria and electrostatic interactions in fluids. Their nonlinear behaviour and mildly alkaline nature reflect intrinsic properties of silicate mineralogies. The sensitivity of fluid pH to soluble elements in low abundance in the host rock, such as carbon, alkali and halogens, illustrates a feedback between the chemistry of the Earth's atmosphere–ocean system^{7,14} and the behavior of metals and volatiles in subduction zones via the composition of the seawater-altered oceanic lithosphere¹⁵. Our results also permit quantification of the electrical conductivity of subduction zone fluids¹¹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 December 2015; accepted 14 September 2016.

- Eugster, H. P. Minerals in hot water. *Am. Mineral.* **71**, 655–673 (1986).
- Hedenquist, J. W. & Lowenstern, J. B. The role of magmas in the formation of hydrothermal ore deposits. *Nature* **370**, 519–527 (1994).
- Pokrovski, G. S. & Dubrovinsky, L. S. The S_3^- ion is stable in geological fluids at elevated temperatures and pressures. *Science* **331**, 1052–1054 (2011).
- Phillips, G. N. & Evans, K. A. Role of CO_2 in the formation of gold deposits. *Nature* **429**, 860–863 (2004).
- Dolejs, D. & Wagner, T. Thermodynamic modeling of non-ideal mineral–fluid equilibria in the system Si–Al–Fe–Mg–Ca–Na–K–H–O–Cl at elevated temperatures and pressures: implications for hydrothermal mass transfer in granitic rocks. *Geochim. Cosmochim. Acta* **72**, 526–553 (2008).
- Evans, K. A. The redox budget of subduction zones. *Earth Sci. Rev.* **113**, 11–32 (2012).
- Hayes, J. M. & Waldbauer, J. R. The carbon cycle and associated redox processes through time. *Phil. Trans. R. Soc. Lond. B* **361**, 931–950 (2006).
- Eiler, J. M., McInnes, B., Valley, J. W., Graham, C. M. & Stolper, E. M. Oxygen isotope evidence for slab-derived fluids in the sub-arc mantle. *Nature* **393**, 777–781 (1998).
- Stolper, E. & Newman, S. The role of water in the petrogenesis of Mariana trough magmas. *Earth Planet. Sci. Lett.* **121**, 293–325 (1994).
- Galvez, M. E. et al. Graphite formation by carbonate reduction during subduction. *Nat. Geosci.* **6**, 473–477 (2013).
- Ni, H., Chen, Q. & Keppler, H. Electrical conductivity measurements of aqueous fluids under pressure with a hydrothermal diamond anvil cell. *Rev. Sci. Instrum.* **85**, 115107 (2014).
- Ding, K. & Seyfried, W. E. Direct pH measurement of NaCl -bearing fluid with an in situ sensor at 400°C and 40 megapascals. *Science* **272**, 1634–1636 (1996).
- Brown, M. Duality of thermal regimes is the distinctive characteristic of plate tectonics since the Neoproterozoic. *Geology* **34**, 961–964 (2006).
- Anbar, A. D. Elements and evolution *Science* **322**, 1481–1483 (2008).
- Staudigel, H. & Hart, S. R. Alteration of basaltic glass: Mechanisms and significance for the oceanic crust–seawater budget. *Geochim. Cosmochim. Acta* **47**, 337–350 (1983).

16. Poli, S. Carbon mobilized at shallow depths in subduction zones by carbonatitic liquids. *Nat. Geosci.* **8**, 633–636 (2015).
17. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
18. Fischer, W. W. *et al.* SQUID–SIMS is a useful approach to uncover primary signals in the Archean sulfur cycle. *Proc. Natl Acad. Sci. USA* **111**, 5468–5473 (2014).
19. Pan, D., Spanu, L., Harrison, B., Sverjensky, D. A. & Galli, G. Dielectric properties of water under extreme conditions and transport of carbonates in the deep Earth. *Proc. Natl Acad. Sci. USA* **110**, 6646–6650 (2013).
20. Sverjensky, D. A., Harrison, B. & Azzolini, D. Water in the deep Earth: the dielectric constant and the solubilities of quartz and corundum to 60 kb and 1200 °C. *Geochim. Cosmochim. Acta* **129**, 125–145 (2014).
21. Galvez, M. E., Manning, C. E., Connolly, J. A. & Rumble, D. The solubility of rocks in metamorphic fluids: a model for rock-dominated conditions to upper mantle pressure and temperature. *Earth Planet. Sci. Lett.* **430**, 486–498 (2015).
22. Mountain, R. D. & Harvey, A. H. Molecular dynamics evaluation of dielectric constant mixing rules for H₂O–CO₂ at geologic conditions. *J. Solution Chem.* **44**, 2179–2193 (2015).
23. Wood, B. J. & Walther, J. V. Rates of hydrothermal reactions. *Science* **222**, 413–415 (1983).
24. Manning, C. E., Antignano, A. & Lin, H. A. Premelting polymerization of crustal and mantle fluids, as indicated by the solubility of albite + paragonite + quartz in H₂O at 1 GPa and 350–620 °C. *Earth Planet. Sci. Lett.* **292**, 325–336 (2010).
25. Aranovich, L. Y. & Newton, R. C. Experimental determination of CO₂–H₂O activity–composition relations at 600–1000 degrees C and 6–14 kbar by reversed decarbonation and dehydration reactions. *Am. Mineral.* **84**, 1319–1332 (1999).
26. Liou, J., Zhang, R. & Ernst, W. Occurrences of hydrous and carbonate phases in ultrahigh-pressure rocks from east-central China: implications for the role of volatiles deep in cold subduction zones. *Isl. Arc* **4**, 362–375 (1995).
27. Rohrbach, A. & Schmidt, M. W. Redox freezing and melting in the Earth's deep mantle resulting from carbon-iron redox coupling. *Nature* **472**, 209–212 (2011).
28. Li, H. & Hermann, J. Apatite as an indicator of fluid salinity: an experimental study of chlorine and fluorine partitioning in subducted sediments. *Geochim. Cosmochim. Acta* **166**, 267–297 (2015).
29. Parkinson, I. J. & Pearce, J. A. Peridotites from the Izu–Bonin–Mariana forearc (ODP Leg 125): evidence for mantle melting and melt–mantle interaction in a supra-subduction zone setting. *J. Petrol.* **39**, 1577–1618 (1998).
30. Mikhail, S. & Sverjensky, D. A. Nitrogen speciation in upper mantle fluids and the origin of Earth's nitrogen-rich atmosphere. *Nat. Geosci.* **7**, 816–819 (2014).
31. Ryabchikov, I., Schreyer, W. & Abraham, K. Compositions of aqueous fluids in equilibrium with pyroxenes and olivines at mantle pressures and temperatures. *Contrib. Mineral. Petrol.* **79**, 80–84 (1982).
32. Wilke, M. *et al.* Zircon solubility and zirconium complexation in H₂O+Na₂O+SiO₂+Al₂O₃ fluids at high pressure and temperature. *Earth Planet. Sci. Lett.* **349–350**, 15–25 (2012).
33. Salvi, S., Pokrovski, G. S. & Schott, J. Experimental investigation of aluminum–silica aqueous complexing at 300 °C. *Chem. Geol.* **151**, 51–67 (1998).

Acknowledgements The presentation of this work benefited from informal reviews by O. Bachmann and D. Rumble. Discussions with P. Ulmer, X. Zhong, J.A. Padron Navarta, D. Miron, D. Sverjensky, J. Eiler and J. Cohen were helpful. This research was supported by an ETH fellowship ETH/CoFUND Fel-06 13-2 (M.E.G.). Partial support from Carnegie and Society in Science/Branco-Weiss fellowships (M.E.G.), Swiss National Science Foundation Grant 200021_146872 (J.A.D.C.), Deep Carbon Observatory and National Science Foundation grant EAR 1347987 (C.E.M.) are also acknowledged.

Author Contributions M.E.G. conceived the project, developed the computational tools used and performed the calculations. J.A.D.C. developed the PerpleX software used for phase equilibria computations. M.E.G., J.A.D.C. and C.E.M. analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.E.G. (matthieu.galvez@erdw.ethz.ch or matthieu.galvez@gmail.com).

Reviewer Information *Nature* thanks D. Dolejs, K. Evans and H. Keppler for their contribution to the peer review of this work.

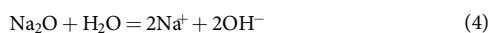
METHODS

Phase equilibria. Phase diagram sections have been computed by forward modeling (Gibbs energy minimization with PerpleX³⁴) using the 2002 revision of the Holland and Powell mineral database³⁵. Rock compositions are chosen to represent model pelite^{36,37}, K-free basalt³⁸ and peridotite (alkali-free lherzolite³⁹) systems (Extended Data Table 1) and maintain H₂O saturation ($a_{\text{H}_2\text{O}} = 1$ in pure H₂O systems, Figs 1, 2a, d) over the entire P – T field for fluid speciation calculation. The carbonated basalt is modified from the C-free composition to maintain fluid saturation (H₂O \approx 5.8 wt%) with \sim 2 wt% CO₂, consistent with average C content of the upper basaltic crust⁴⁰. Note that a harzburgite peridotite composition (Fe–Mg–Al–Si system) would dampen the ionic strength and Δ pH of the equilibrium fluid without significant change to their P – T behaviour. Component \mathcal{F} is modified from a representative crustal fluid composition from ref. 41. The solid solution models used include biotite (Bio)⁴², garnet (Grt)³⁵, chlorite (Chl)⁴³, staurolite (Sta)⁴⁴, white mica (Mus, Phe or Par)^{45,46}, K-feldspars (Kfs)⁴⁷, clinopyroxene (Cpx)⁴⁸, cordierite (Crđ)⁴⁴, chloritoid (Ctd)⁴⁹, orthopyroxene (Opx)⁴⁸, plagioclase (Pl)⁵⁰, amphibole (Amp)⁵¹, carbonate (Do)³⁵, magnesite³⁵, antigorite (Atg)³⁹, olivine (O)³⁵ and an ideal solution model for talc (T). Minerals considered pure are quartz (Q), lawsonite (Law), kyanite (Ky), zoisite (Zo) and sillimanite (Sill). Additional mineral abbreviations: albite (Ab), jadeite (Jd), paragonite (Par), brucite (Brc). Two variants of the metasomatic model were considered: with (Fig. 3c, d) and without Na-phlogopite (aspidolite⁵²) for which thermochemical data are uncertain. In the absence of this phase, Δ pH (where $X_{\mathcal{F}} = 0.01$ – 0.43) increases by \sim 0.5 unit without change to the main features of the diagram. Wet melting curves for basalt and pelite are indicative, and drawn from experimental data in refs 53–55, carbonated-basalt melting curve is from experimental data in ref. 16.

P – T paths. Two model top-of-the-slab P – T paths represent cold (Central-Honshu⁵⁶ referred to as ‘Honshu’) and intermediate (Columbia-Ecuador⁵⁶ referred to as ‘Columbia’) thermal regimes. An indicative hot P – T path representative of Archaean subduction has been computed with a mantle potential temperature of $T = 1,650^\circ\text{C}$ and parameters from the Cascadia subduction zone⁵⁶. We thank P. Van Keken for computing and providing the P – T coordinates of this geotherm. In combination, these paths are reasonable approximations for a broad range of possible geotherms likely to be realized in subduction zones from the Archaean to the Phanerozoic⁵⁷.

Solvent definition. Neutrally charged solvent species (for example, H₂O, CO₂, CH₄) are described by non-ideal molecular solution models. They are used to compute the phase diagram sections (Fig. 1) and the solvent dielectric constant (see below). Two distinct models are used for C-saturated C–O–H solvents⁵⁸ (graphitic pelite in Fig. 2b), and for binary H₂O–CO₂ solvents⁵⁹ (carbonated basalt in Fig. 2b). They define the solvents in which solute speciation is computed. This strategy differs from previous approaches^{60,61} in that C solubility predictions are consistent with independent constraints from both low- T congruent mineral dissolution (calcite⁶²) and high- T mixed-volatile carbonate-silicate incongruent dissolution experiments²⁵ over an extensive range of P , T and water activities. Our terminology differs in that congruent and incongruent carbonate dissolution at sub-solidus conditions are distinguished here, as opposed to dissolution versus decarbonation elsewhere^{61,63}.

Speciation calculation. At equilibrium the oxide chemical potentials are linked to the partial molar Gibbs energy of corresponding dissociated species through general hydrolysis equilibria such as:



This equation, and equivalent ones for all solutes considered (Extended Data Table 2) generates a set of equations of the form presented in Equation (1) (see main text). Together with water hydrolysis:



charge balance in the fluid $\sum_j Z_j m_j = 0$ (Z_j is the ionic charge, m_j the molality) and

an activity model for electrolytes, the equations governing fluid speciation are closed.

The chemical potentials of oxide components are obtained by differentiation of the extensive Gibbs energy of the system. It is assumed that the solutes do not affect the rock composition and mineralogy. This is a reasonable assumption in high- P subduction zone environments characterized by elevated rock/fluid ratio of interest here. The standard-state Gibbs energy of aqueous species is computed with the Helgeson–Kirkham–Flowers equation of state (EoS)^{20,64–68} using the Born model, extended here to mixed-volatile carbonic solvents²¹, adopting the convention $g_{\text{H}^+}^\ominus = 0$ (see main text). The thermochemical database for aqueous species and sources of those data are compiled in Extended Data Table 2. Consistency between

mineral and solute thermochemical database is judged adequate for the present purpose. Further discussions can be found in refs 21, 35, 69, 70.

Dissolved solutes that exert a dominant control on charge balance, ionic strength and computed pH, are Na⁺, OH[−] and HSiO₃[−] at all conditions for the pelite and basalt fluids, and Ca²⁺, CaOH⁺ and OH[−] (and AlO₂[−] along the hot geotherm) for the peridotite fluids. We estimate the precision on absolute Δ pH values is ± 0.2 Δ pH (mainly due to uncertainty of thermodynamic data). In absence of experimental validations, the predictive accuracy is within 1 Δ pH unit. Relative variations and discontinuities are robust.

Fluid dielectric properties, standard state and activity model for solutes. The dielectric constant (static relative permittivity) of water is from refs 19, 20, 71 and 72, used in the Born dielectric continuum model of electrolyte solvation⁷³. Those of the mixed-volatile solvents are computed with a method developed in ref. 21. It uses the Landau–Lifschitz–Looyenga mixing scheme^{22,74,75} with volume fractions of solvent components computed with a reduced density method⁷⁶. Input solvent density at critical and target P , T conditions are derived from the molecular fluid EoS (see above).

The energetics of the solutes in going from pure H₂O to a mixed-volatile solvent characterized by ϵ_{mix} , is incorporated to the standard-state Gibbs energy of the solute^{77,78}, that is, it affects the equilibrium constants of homogeneous oxide-solutes equilibria in solution. Solute standard-state is unit activity of 1 mol of solute in 1 kg of solvent, referenced to infinite dilution. Aqueous species molal concentration is given in mol per kg of solvent.

Non-ideality for electrolytes follows the Davies extension of the Debye–Hückel law in its original 1938 formulation⁷⁹. The ionic strength of the solution (I) is linked to the activity coefficient (γ_j) of the j th anion or cation by $\log \gamma_j = \frac{-AZ_j^2 I^{1/2}}{(1 + I^{1/2})} + 0.2Z_j^2 I$,

where $I = \frac{1}{2} \sum_j Z_j^2 m_j$, $A = 1.82483 \times 10^6 \rho_{\text{mix}}^{1/2} (\epsilon_{\text{mix}} T)^{-3/2}$, m_j refers to the mola-

lity of the ion, A is the Debye–Hückel parameter, and ρ_{mix} and ϵ_{mix} are respectively the density and dielectric constant of the solvent. Using a coefficient $C = 0.3$ (ref. 79) instead of $C = 0.2$ (ref. 80) in the empirical term $CZ_j^2 I$ introduced no noticeable change to the results presented. The activity coefficient for neutral solutes is 1. The concentration of chlorine (for example, Fig. 2c, f) is set by fixing the total chlorine (ΣCl) content of the fluid (Cl[−], HCl, NaCl \pm KCl solutes) to $\Sigma \text{Cl} = 1$ molal. Further technical details can be found in ref. 21.

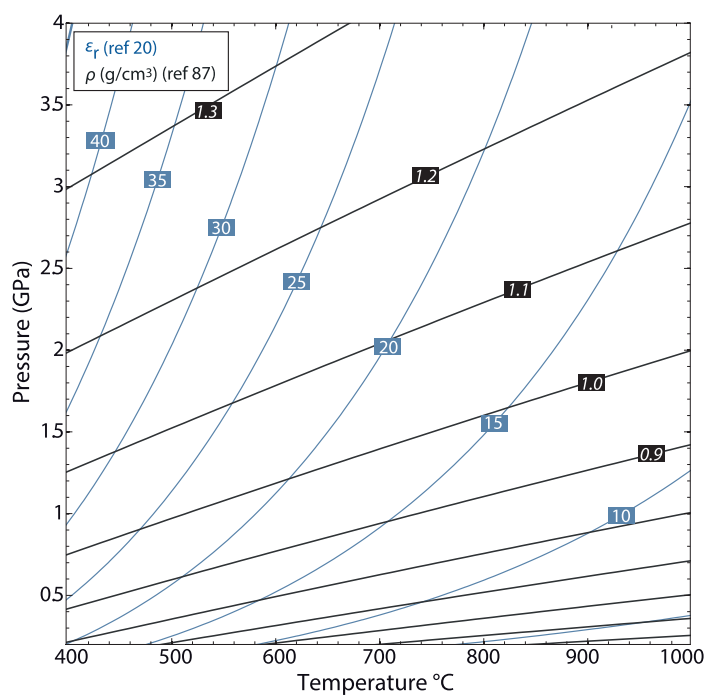
C flux estimates along cold and hot paths. Although subduction zone elemental budgets are sensitive to lithosphere hydration, overall length of subduction trenches, composition of subducting materials, geotherms and thus fluid fluxes, provisional estimates can be derived combining element solubility (this work) with fore-arc (<100 km depth, ref 81) and sub-arc (100 to 150 km depth) H₂O flux computed from ref. 81, assuming 2% H₂O in the upper 4 km of the oceanic mantle (Extended Data Table 3). Focusing on carbon, assuming a total length of 3.85×10^7 m of subduction trenches, a cold subduction (model based on the central Honshu system) results in a global annual sub-solidus C flux, combining forearc and sub-arc fluxes, of $\phi_{\text{C}} = 10 \pm 7$ Mt yr^{−1}. This corresponds to a net equivalent O flux beyond subarc depth of about 105 ± 50 Mt yr^{−1} (or redox budget flux (see ref 6) of 14 ± 7 Tmol yr^{−1} assuming C⁰ (elemental C) is the reference C oxidation state in the deep upper mantle), assuming a subduction C input of 40 – 66 Mt yr^{−1} (refs 61). Conversely, along the hot geotherm (Figs 1, 3b), the forearc flux alone is $\phi_{\text{C}} = 40 \pm 7$ Mt yr^{−1} that is within uncertainty of present-day subduction C input in ref. 61. This budget contrasts with cold regimes where up to 85% of the C input is retained in slabs, in agreement with conclusions from refs 82–85, and in partial contrast with challenging views⁶¹. Hot and intermediate paths (Columbia, Fig. 1a) typical of the Precambrian and rare modern settings are the only scenarios where substantial, possibly quantitative, forearc decarbonation may occur. In the latter intermediate scenario, quantitative C loss may occur, yet not by sub-solidus dissolution, but by carbonate melting^{16,83,86}. Slabs retaining \sim 40% of their initial C budget would cross the carbonated-basalt melting curve at sub-arc depth (Fig. 1a, see refs 16, 83, 86). Note that this temporal trend is consistent with previous interpretation of kimberlite–diamond isotopic signatures reflecting little recycled components in the mantle⁶¹. The latter isotopic signatures may record C recycling under dominantly hot Archaean and proterozoic subduction thermal regimes, and may not be representative of Phanerozoic and present-day subduction zones.

34. Connolly, J. A. D. Multivariable phase diagrams; an algorithm based on generalized thermodynamics. *Am. J. Sci.* **290**, 666–718 (1990).

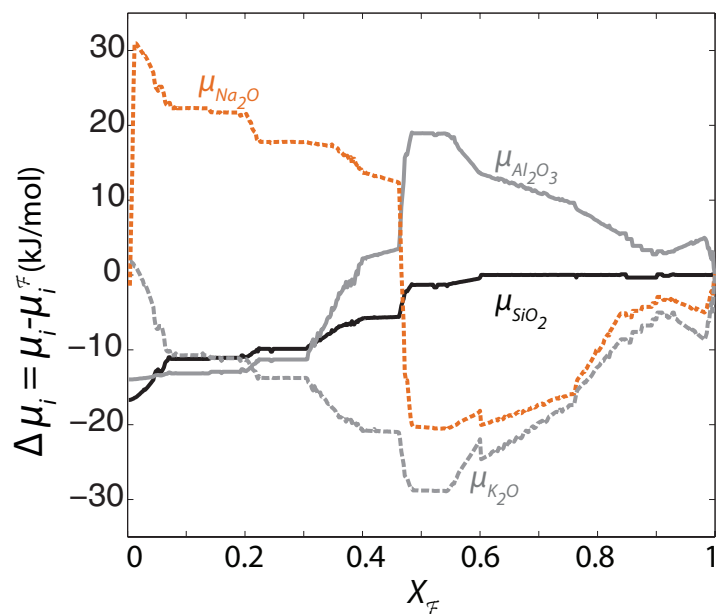
35. Holland, T. & Powell, R. An internally consistent thermodynamic data set for phases of petrological interest. *J. Metamorph. Geol.* **16**, 309–343 (1998).

36. Caddick, M. J. & Thompson, A. B. Quantifying the tectono-metamorphic evolution of pelitic rocks from a wide range of tectonic settings: mineral compositions in equilibrium. *Contrib. Mineral. Petrol.* **156**, 177–195 (2008).

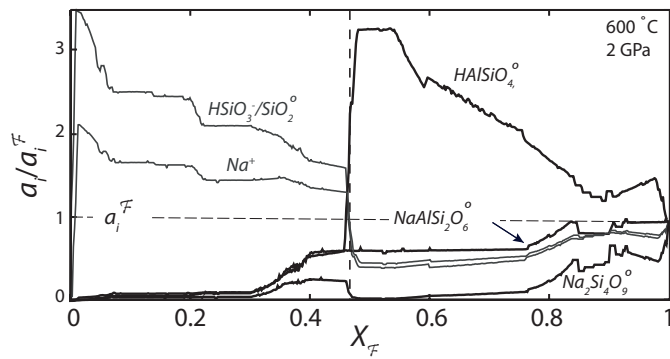
37. Shaw, D. M. Geochemistry of pelitic rocks. Part III: Major elements and general geochemistry. *Geol. Soc. Am. Bull.* **67**, 919–934 (1956).
38. Pearce, J. A. Statistical analysis of major element patterns in basalts. *J. Petrol.* **17**, 15–43 (1976).
39. Padrón-Navarta, J. A. *et al.* Tschermak's substitution in antigorite and consequences for phase relations and water liberation in high-grade serpentinites. *Lithos* **178**, 186–196 (2013).
40. Alt, J. C. & Teagle, D. A. H. The uptake of carbon during alteration of ocean crust. *Geochim. Cosmochim. Acta* **63**, 1527–1535 (1999).
41. Hermann, J., Zheng, Y.-F. & Rubatto, D. Deep fluids in subducted continental crust. *Elements* **9**, 281–287 (2013).
42. Tajčmanová, L., Connolly, J. A. D. & Cesare, B. A thermodynamic model for titanium and ferric iron solution in biotite. *J. Metamorph. Geol.* **27**, 153–165 (2009).
43. Holland, T., Baker, J. & Powell, R. Mixing properties and activity-composition and relationships of chlorites in the system $\text{MgO-FeO-Al}_2\text{O}_3\text{-SiO}_2\text{-H}_2\text{O}$. *Eur. J. Mineral.* **10**, 395–406 (1998).
44. Mahar, E. M., Baker, J. M., Powell, R., Holland, T. J. B. & Howell, N. The effect of Mn on mineral stability in metapelites. *J. Metamorph. Geol.* **15**, 223–238 (1997).
45. Coggon, R. & Holland, T. J. B. Mixing properties of phengitic micas and revised garnet-phengite thermobarometers. *J. Metamorph. Geol.* **20**, 683–696 (2002).
46. Auzanneau, E., Schmidt, M. W., Vielzeuf, D. & Connolly, J. D. Titanium in phengite: a geobarometer for high temperature eclogites. *Contrib. Mineral. Petrol.* **159**, 1–24 (2010).
47. Waldbaum, D. R. & Thompson, J. B. Mixing properties of sanidine crystalline solutions. 2. Calculations based on volume data. *Am. Mineral.* **53**, 2000–2017 (1968).
48. Holland, T. & Powell, R. Thermodynamics of order-disorder in minerals: II. Symmetric formalism applied to solid solutions. *Am. Mineral.* **81**, 1425–1437 (1996).
49. White, R. W., Powell, R., Holland, T. J. B. & Worley, B. A. The effect of TiO_2 and Fe_2O_3 on metapelitic assemblages at greenschist and amphibolite facies conditions: mineral equilibria calculations in the system $\text{K}_2\text{O-FeO-MgO-Al}_2\text{O}_3\text{-SiO}_2\text{-H}_2\text{O-TiO}_2\text{-Fe}_2\text{O}_3$. *J. Metamorph. Geol.* **18**, 497–511 (2000).
50. Newton, R. C., Charlu, T. V. & Kleppa, O. J. Thermochemistry of the high structural state plagioclases. *Geochim. Cosmochim. Acta* **44**, 933–941 (1980).
51. Dale, J., Powell, R., White, R., Elmer, F. & Holland, T. A thermodynamic model for Ca-Na clin amphiboles in $\text{Na}_2\text{O-CaO-FeO-MgO-Al}_2\text{O}_3\text{-SiO}_2\text{-H}_2\text{O-O}$ for petrological calculations. *J. Metamorph. Geol.* **23**, 771–791 (2005).
52. Pirard, C. & Hermann, J. Experimentally determined stability of alkali amphibole in metasomatised dunite at sub-arc pressures. *Contrib. Mineral. Petrol.* **169**, 1–26 (2015).
53. Hack, A. C., Thompson, A. B. & Aerts, M. Phase relations involving hydrous silicate melts, aqueous fluids, and minerals. *Rev. Mineral. Geochem.* **65**, 129–185 (2007).
54. Kessel, R., Ulmer, P., Pettke, T., Schmidt, M. W. & Thompson, A. B. The water-basalt system at 4 to 6 GPa: phase relations and second critical endpoint in a K-free eclogite at 700 to 1400°C. *Earth Planet. Sci. Lett.* **237**, 873–892 (2005).
55. Schmidt, M. W. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci. Lett.* **163**, 361–379 (1998).
56. Syracuse, E. M., van Keken, P. E. & Abers, G. A. The global range of subduction zone thermal models. *Phys. Earth Planet. Inter.* **183**, 73–90 (2010).
57. Penniston-Dorland, S. C., Kohn, M. J. & Manning, C. E. The global range of subduction zone thermal structures from exhumed blueschists and eclogites: rocks are hotter than models. *Earth Planet. Sci. Lett.* **428**, 243–254 (2015).
58. Connolly, J. A. D. & Cesare, B. C-O-H-S fluid composition and oxygen fugacity in graphitic metapelites. *J. Metamorph. Geol.* **11**, 379–388 (1993).
59. Holland, T. & Powell, R. Activity-composition relations for phases in petrological calculations: an asymmetric multicomponent formulation. *Contrib. Mineral. Petrol.* **145**, 492–501 (2003).
60. Kerrick, D. M. & Connolly, J. A. D. Metamorphic devolatilization of subducted marine sediments and the transport of volatiles into the Earth's mantle. *Nature* **411**, 293–296 (2001).
61. Kelemen, P. B. & Manning, C. E. Reevaluating carbon fluxes in subduction zones, what goes down, mostly comes up. *Proc. Natl Acad. Sci. USA* **112**, E3997–E4006 (2015).
62. Caciagli, N. C. & Manning, C. E. The solubility of calcite in water at 6–16 kbar and 500–800°C. *Contrib. Mineral. Petrol.* **146**, 275–285 (2003).
63. Frezzotti, M. L., Selverstone, J., Sharp, Z. D. & Compagnoni, R. Carbonate dissolution during subduction revealed by diamond-bearing rocks from the Alps. *Nat. Geosci.* **4**, 703–706 (2011).
64. Helgeson, H. C., Kirkham, D. H. & Flowers, G. C. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes by high pressures and temperatures; IV, Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600 degrees C and 5 kb. *Am. J. Sci.* **281**, 1249–1516 (1981).
65. Tanger, J. C. & Helgeson, H. C. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures; revised equations of state for the standard partial molal properties of ions and electrolytes. *Am. J. Sci.* **288**, 19–98 (1988).
66. Helgeson, H. C. & Kirkham, D. H. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures; II, Debye-Huckel parameters for activity coefficients and relative partial molal properties. *Am. J. Sci.* **274**, 1199–1261 (1974).
67. Helgeson, H. C. & Kirkham, D. H. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures; I, Summary of the thermodynamic/electrostatic properties of the solvent. *Am. J. Sci.* **274**, 1089–1198 (1974).
68. Helgeson, H. C. & Kirkham, D. H. Theoretical prediction of the thermodynamic properties of aqueous electrolytes at high pressures and temperatures. III. Equation of state for aqueous species at infinite dilution. *Am. J. Sci.* **276**, 97–240 (1976).
69. Sverjensky, D. A., Hemley, J. & d'Angelo, W. Thermodynamic assessment of hydrothermal alkali feldspar-mica-aluminosilicate equilibria. *Geochim. Cosmochim. Acta* **55**, 989–1004 (1991).
70. Miron, G. D., Wagner, T., Kulik, D. A. & Heinrich, C. A. Internally consistent thermodynamic data for aqueous species in the system Na-K-Al-Si-O-H-Cl. *Geochim. Cosmochim. Acta* **187**, 41–78 (2016).
71. Fernández, D. P., Mulev, Y., Goodwin, A. & Sengers, J. L. A database for the static dielectric constant of water and steam. *J. Phys. Chem. Ref. Data* **24**, 33–70 (1995).
72. Fernández, D., Goodwin, A., Lemmon, E. W., Sengers, J. L. & Williams, R. A formulation for the static permittivity of water and steam at temperatures from 238 K to 873 K at pressures up to 1200 MPa, including derivatives and Debye-Hückel coefficients. *J. Phys. Chem. Ref. Data* **26**, 1125–1166 (1997).
73. Bockris, J. O. M. & Reddy, A. K. N. *Modern Electrochemistry: An Introduction to an Interdisciplinary Area* Vol. 2 (Springer, 1973).
74. Landau, L. D. & Lifshitz, M. *Electrodynamics of Continuous Media* (Pergamon, 1960).
75. Looyenga, H. Dielectric constants of heterogeneous mixtures. *Physica* **31**, 401–406 (1965).
76. Harvey, A. H. & Prausnitz, J. M. Dielectric constants of fluid mixtures over a wide range of temperature and density. *J. Solution Chem.* **16**, 857–869 (1987).
77. Walther, J. V. Ionic association in $\text{H}_2\text{O-CO}_2$ fluids at mid-crustal conditions. *J. Metamorph. Geol.* **10**, 789–797 (1992).
78. Akinfiev, N. & Zotov, A. Thermodynamic description of equilibria in mixed fluids (H_2O -non-polar gas) over a wide range of temperature (25–700°C) and pressure (1–5000 bars). *Geochim. Cosmochim. Acta* **63**, 2025–2041 (1999).
79. Davies, C. W. *Ion Association* (Butterworths, 1962).
80. Davies, C. W. The extent of dissociation of salts in water. Part VIII. An equation for the mean ionic activity coefficient of an electrolyte in water, and a revision of the dissociation constants of some sulphates. *J. Chem. Soc.* 2093–2098 (1938).
81. van Keken, P. E., Hacker, B. R., Syracuse, E. M. & Abers, G. A. Subduction factory: 4. Depth-dependent flux of H_2O from subducting slabs worldwide. *J. Geophys. Res. Solid Earth* **116**, B01401 (2011).
82. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298**, 1–13 (2010).
83. Thomson, A. R., Walter, M. J., Kohn, S. C. & Brooker, R. A. Slab melting as a barrier to deep carbon subduction. *Nature* **529**, 76–79 (2016).
84. Connolly, J. A. D. Computation of phase equilibria by linear programming: a tool for geodynamic modeling and its application to subduction zone decarbonation. *Earth Planet. Sci. Lett.* **236**, 524–541 (2005).
85. Gorman, P. J., Kerrick, D. M. & Connolly, J. A. D. Modeling open system metamorphic decarbonation of subducting slabs. *Geochem. Geophys. Geosyst.* **7**, Q04007 (2006).
86. Skora, S. *et al.* Hydrous phase relations and trace element partitioning behaviour in calcareous sediments at subduction-zone conditions. *J. Petrol.* **56**, 953–980 (2015).
87. Zhang, Z. & Duan, Z. Prediction of the PVT properties of water over wide range of temperatures and pressures from molecular dynamics simulation. *Phys. Earth Planet. Inter.* **149**, 335–354 (2005).
88. Pokrovskii, V. A. & Helgeson, H. C. Thermodynamic properties of aqueous species and the solubilities of minerals at high pressures and temperatures; the system $\text{Al}_2\text{O}_3\text{-H}_2\text{O-NaCl}$. *Am. J. Sci.* **295**, 1255–1342 (1995).
89. Pokrovskii, V. A. & Helgeson, H. C. Thermodynamic properties of aqueous species and the solubilities of minerals at high pressures and temperatures; the system $\text{Al}_2\text{O}_3\text{-H}_2\text{O-KOH}$. *Chem. Geol.* **137**, 221–242 (1997).
90. Sverjensky, D. A., Shock, E. L. & Helgeson, H. C. Prediction of the thermodynamic properties of aqueous metal complexes to 1000°C and 5 kb. *Geochim. Cosmochim. Acta* **61**, 1359–1412 (1997).
91. Xie, Z. & Walther, J. V. Wollastonite+ quartz solubility in supercritical NaCl aqueous solutions. *Am. J. Sci.* **293**, 235–255 (1993).
92. Miron, G. D. *Internally Consistent Thermodynamic Database for Fluid-Rock Interaction: Tools, Methods and Optimization*. Dissertation no. 23242, ETH-Zurich (2016).



Extended Data Figure 1 | Properties of water at geological conditions. The relative permittivity²⁰ (ϵ_r) and density⁸⁷ (ρ) of pure water as functions of pressure and temperature. Adapted from ref. 21.



Extended Data Figure 2 | Oxide chemical potential across a peridotite–crust interface. Shown are profiles of oxide chemical potentials (μ) in the metasomatic model (Fig. 3c, d) relative to that for component F (μ_i^F) at 600 °C and 2 GPa.



Extended Data Figure 3 | Relative activity of selected neutral species across a serpentinite–crust interface. Shown are relative activities

$$\left(\frac{a_i}{a_i^F}\right)_{P,T}^{\text{fluid}} = \exp \left[\frac{1}{RT} \left(\mu_i - \mu_i^F \right)_{P,T}^{\text{rock}} \right]$$

of neutral polynuclear clusters HAlSiO_4^0 , $\text{NaAlSi}_2\text{O}_6^0$ and $\text{Na}_2\text{Si}_4\text{O}_9^0$, as well as Na^+ and $\text{HSiO}_3^-/\text{SiO}_2^0$, in the fluid, at 600 °C and 2 GPa (see Fig. 3c, d).

Extended Data Table 1 | Rock compositions used for phase equilibria computations

	SiO ₂	Al ₂ O ₃	Na ₂ O	K ₂ O	MgO	CaO	FeO	H ₂ O	CO ₂
Pelite ⁱ	65.6	18.17	1.89	3.87	2.87	1.19	6.37	sat. ^{vi}	
Basalt ⁱⁱ	51.6	16	2.6	0	7.7	11.3	10.8	sat.	
Basalt+C ⁱⁱⁱ	47.5	14.7	2.4	0	7.1	10.4	9.9	5.85	2.02
Peridotite ^{iv}	46.11	4.52	0	0	39.05	3.26	7.04	sat.	
Fcomponent ^v	70.11	12.74	5.16	2.61	2.24	3.11	3.99	sat.	

^aComposition from ref. 36, the C-saturated pelite with fluid atomic ratio O/H = 0.5 has a similar oxide composition.

^bModified from ref. 38.

^cModified from ref. 38.

^dComposition from ref. 39.

^eComposition from ref. 41.

^fsat., saturated.

Extended Data Table 2 | Thermodynamic data source for solutes used in this study

Inorganic species				C species				Cl species	
SiO_2^0	<i>a</i>	KOH^0	<i>d</i>	MgSiO_3^0	<i>h</i>	HCO_3^-	<i>a</i>	HCl^0	<i>a</i>
Si_2O_4^0	<i>a</i>	K^+	<i>a</i>	Fe^{2+}	<i>a</i>	CO_3^{2-}	<i>a</i>	NaCl^0	<i>a</i>
HSiO_3^-	<i>b</i>	KAlO_2^0	<i>e</i>	OH^-	<i>a</i>	HCOO^-	<i>a</i>	KCl^0	<i>a</i>
HAlO_2^0	<i>a</i>	Ca(OH)^+	<i>a</i>			NaCO_3^-	<i>a</i>	Cl^-	<i>a</i>
AlO_2^-	<i>a</i>	Ca^{2+}	<i>a</i>			NaHCO_3^0	<i>a</i>		
Na(OH)^0	<i>a</i>	CaHSiO_3^+	<i>f</i>			CaHCO_3^+	<i>a</i>		
Na^+	<i>a</i>	CaSiO_3^0	<i>g</i>			MgHCO_3^+	<i>a</i>		
NaAlO_2^0	<i>c</i>	Mg(OH)^+	<i>a</i>			MgCO_3^0	<i>a</i>		
NaHSiO_3^0	<i>a</i>	MgHSiO_3^+	<i>a</i>			CaCO_3^0	<i>a</i>		

^aData as reported in the Deep Earth Water (DEW) database, (ref. 20).^bData as reported in the DEW database, (ref. 20).^cFrom ref. 88.^dAs reported in ref. 20.^eFrom ref. 89.^fModified from ref. 90 to obtain a_1 parameter (HKF volumetric coefficient) consistent with the sum of the revised a_1 of Ca^{2+} and HSiO_3^- .^gEstimates from data at 25 °C for CaSiO_3^0 from cement phase solubility, and wollastonite solubility from ref. 91 (compare ref. 92).^hAssumed to have the same logK of formation as CaSiO_3^0 , from ref. 92.

Extended Data Table 3 | Estimates of global annual C fluxes

	H ₂ O flux ^a (Mt/yr)		C concentration (mol/kg H ₂ O)		C flux (Mt/yr)		
	forearc	subarc ^b	forearc	subarc	forearc	subarc	total
Central Honshu	277	177	0.3–3 ^c	1–3 ^d	5 ± 4	4 ± 2	10 ± 7
Columbia-Ecuador	420	454	1–10	7–10 [*]	28 ± 23	46 ± 8 [*]	74 ± 31 [*]
Mexico ('Archean') ^f	474	235	4–10 ^{e,*}	7–10 [*]	40 ± 17 [*]	24 ± 4 [*]	64 ± 21 [*]
Global average	331	130	0.3–6 [*]	3–6 [*]	13 ± 11 [*]	7 ± 2 [*]	20 ± 14 [*]

Shown are estimates of global annual C fluxes derived from C solubility data (this study) and H₂O fluxes from subducting slabs (ref. 81).

^aThe global annual H₂O flux for each subduction locality is obtained by multiplication of the annual H₂O flux per metre of trench by the entire length of the present subduction trenches, that is, 38.5×10^6 m (ref. 81). The global average is obtained from ref. 81. C concentrations for the global (present-day) average are indicative, intermediate between the Honshu and Columbia endmembers scenario.

^bSubarc H₂O flux are released from slabs between 100 and 150 km depth.

^cThe value of 0.3 mol C per kg H₂O is at ~500 °C (Fig. 3a), and 3 mol C per kg H₂O is a conservative estimate at ~650–700 °C and ~3 GPa (Fig. 3a).

^dThe concentration range 1–3 mol C per kg H₂O accounts for the decrease in C solubility with *P* increase along the high-T section of the P–T path.

^eThe value of 4 mol C per kg H₂O is at 600 °C, Fig. 3b, increasing to ~10 mol C per kg H₂O at ~700–750 °C.

^fThe H₂O fluxes for the model Archean scenario are those computed from the Mexico geotherm (ref. 81).

^{*}The solidus may be crossed along the path. Rock partial melting may significantly affect the values reported.

LETTER

doi:10.1038/nature20140

Variability in plant nutrients reduces insect herbivore performance

William C. Wetzel, Heather M. Kharouba, Moria Robinson, Marcel Holyoak & Richard Karban

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. *Nature* is providing this early version of the typeset paper as a service to our customers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Cite this article as: Wetzel, W. *et al.* Variability in plant nutrients reduces insect herbivore performance. *Nature* <http://dx.doi.org/10.1038/nature20140> (2016).

Received 29 November 2015; accepted 5 October 2016.
Accelerated Article Preview Published online 12 October 2016.

Variability in plant nutrients reduces insect herbivore performance

William C. Wetzel^{1,2}, Heather M. Kharouba^{2,3}, Moria Robinson², Marcel Holyoak⁴ & Richard Karban⁵

The performance and population dynamics of insect herbivores depend on the nutritive and defensive traits of their host plants¹. The literature on plant-herbivore interactions focuses on plant trait means^{2–4}, but recent studies showing the importance of plant genetic diversity for herbivores suggest that plant trait variance may be equally important^{5,6}. The consequences of plant trait variance for herbivore performance, however, have been largely overlooked. Here we report an extensive assessment of the effects of within-population plant trait variance on herbivore performance using 457 performance datasets from 53 species of insect herbivores. We found that variance in plant nutritive traits substantially reduces mean herbivore performance via nonlinear averaging of performance relationships that were overwhelmingly concave-down. In contrast, relationships between herbivore performance and plant defense levels were typically linear, such that plant defense variance does not affect herbivore performance via nonlinear averaging. Our results demonstrate that plants contribute to the suppression of herbivore populations by having variable nutrient levels, not just by having low average quality as is typically thought. We propose that this phenomenon could play a key role in the suppression of herbivore populations in natural systems, and that increased nutrient heterogeneity within agricultural crops could contribute to the sustainable control of insect pests in agroecosystems.

Decades of research have established the importance of plant nutritive and defensive traits for herbivore performance and population dynamics¹. Recent studies, showing that plant genetic diversity influences herbivore community patterns, suggest that plants influence herbivores not just through average trait values but also through variance in trait values^{5,6}. The literature on plant defenses and herbivore nutritional ecology, however, focuses on mean relationships and mostly ignores the consequences of trait variance^{2,3}. This is an oversight because intraspecific plant trait variance pervades natural systems, from among tissues within individuals to among individuals within populations⁴. In modern agroecosystems, however, plant functional diversity has been replaced by extensive homogeneous monocultures of single crop varieties or genotypes⁷. How the loss of trait diversity influences higher trophic levels and ecosystems services like pest control is unexplored relative to how much is known about the consequences of genetic diversity^{8,9}. Elucidation of the direct effects of variability in plant defensive or nutritive traits on herbivore performance would inform management of agroecosystems—perhaps revealing new ways to use crop heterogeneity for sustainable pest management—and advance our fundamental understanding of plant-insect interactions. Here we test for general patterns in the effects of plant trait variance on herbivore performance using 457 datasets relating plant traits to herbivore growth and survival for 53 species of phytophagous insects from seven orders.

Plant variance could influence herbivores in several ways, including reducing the opportunity for herbivore populations to adapt

evolutionarily to plant defenses¹⁰. We focused on the ecological effects that occur via nonlinear averaging, a general phenomenon that potentially applies to all species and has been used successfully to understand diverse biological phenomena^{11,12}. In this context, nonlinear averaging—also known as Jensen's inequality¹³—allows us to predict the effects of plant variance on mean herbivore performance using the curvature of the relationship between plant trait values and herbivore performance (Fig. 1). If the function relating herbivore performance to a plant trait is concave-down (decelerating), then herbivore populations experiencing variance in that trait will have lower mean performance relative to herbivores experiencing constant levels of the trait at the same trait mean³. In contrast, variance enhances mean performance when herbivore performance functions are concave-up (accelerating). When performance functions are linear, plant trait variance has no effect. Three reviews have attempted to generalize the effects of plant trait variance on herbivore performance by visually assessing and categorizing the curvature of herbivore performance functions in published studies^{3,4,11}. They came to contradictory conclusions, perhaps because they had small sample sizes (< 12 studies) and lacked objective methods for quantifying curvature.

We found 76 papers published between 1968 and 2014 that allowed us to estimate herbivore performance functions (Supplementary Tables 1 and 2; Extended Data Fig. 1). These papers reported growth or survival of herbivores at ≥ 4 values of a plant trait. Our search only included studies that directly manipulated trait values in a laboratory setting (e.g., via artificial diet, or applying compounds to plant surfaces) to avoid inclusion of spurious correlations. We estimated a performance function for each dataset with a cubic spline (e.g., Fig. 2, A and B). The spline from each dataset allowed us to quantify the effect of trait variance on herbivore performance, specifically the difference in performance in the presence and absence of plant trait variance (Jensen's effect). We predicted performance in the absence of variance using the value of the performance function at the mean of the plant trait levels in the dataset. We predicted performance in the presence of variance using the mean of the values of the performance function at each of the trait levels (Extended Data Fig. 2). This approach assumes the original authors chose trait levels reflective of trait distributions encountered by herbivore populations in nature. Indeed, many studies reported field data justifying their range of values. In nature the magnitude—but not the sign—of Jensen's effect will depend on the plant trait distribution and resource selection behavior; however, even herbivore populations with highly selective resource behavior will experience plant variance due to costs and limits of discrimination and intraspecific competition⁴. We tested sensitivity of our conclusions to the shape of the trait distribution by repeating the entire analysis assuming uniform and Gaussian distributions.

First, we asked how the consequences of plant trait variance for herbivore performance differed for plant defensive traits (e.g., concentrations of toxic secondary metabolites) versus plant nutritive traits

¹Department of Entomology and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA. ²Center for Population Biology, University of California, Davis, Davis, CA, USA.

³Department of Biology, University of Ottawa, Ottawa, ON, Canada. ⁴Department of Environmental Science and Policy, University of California, Davis, Davis, CA, USA. ⁵Department of Entomology and Nematology, University of California, Davis, Davis, CA, USA.

(e.g., protein concentration). Physiological theory predicts the relationship between nutritive traits and herbivore performance should be concave-down^{14–17} (Fig. 2D), but makes less clear predictions about the relationship between defensive traits and performance^{18,19} (Fig. 2C). Despite the centrality of this question to insect physiology and broader theory in plant-insect interactions, we know of no other quantitative analyses of shapes of empirical performance curves.

We found that relationships between plant nutrients and herbivore growth and survival were consistently concave down, whereas relationships between plant defenses and herbivore growth and survival were close to linear on average. The curvature of nutrient-performance relationships led the experimentally-generated variability in nutrients to reduce mean herbivore growth by half a standard deviation relative to plants with a similar mean nutrient value but zero nutrient variance (Jensen's effect [95% CI] = -0.49 [-0.62, -0.35]) (Fig. 3A). The mean effect of nutrient variance was more than half a standard deviation more negative (-0.55 [-0.66, -0.44]) than the mean effect of defense variance, which was near zero (0.063 [-0.055, 0.18]; $\chi^2_1 = 92.6$, $P < 0.0001$). This difference suggests there is less influence of variance in plant defenses on herbivore performance compared to variance in nutritional content. Mirroring the results for growth variables, studies of herbivore survival had negative Jensen's effects for nutritive traits (-1.41 [-2.29, -0.53]) and effects near zero (linear) for defensive traits (0.28 [-0.24, 0.81]; $\chi^2_1 = 19.2$, $P < 0.0001$) (Fig. 3B).

These results indicate that there are consistent constraints on herbivore physiology that lead to concave-down nutrient-performance relationships (Fig. 3, D and F) and depressed growth and survival in the face of plant nutrient variability (Fig. 3 A and B). This finding supports the theoretical prediction that performance for most consumers increases with nutrients but then plateaus due to diminishing returns, or even declines at high nutrient levels due to nutrient toxicity^{14,17}. In contrast, the generally linear declines in performance with increasing levels of plant defense indicate that defense variability has little effect on herbivores via nonlinear averaging (Fig. 3, C and E). This finding contradicts recent predictions about the ubiquity of hormesis—beneficial, stimulatory effects of low doses of toxins¹⁸—or indicates that hormesis may occur only at lower doses than were tested by studies in our sample. It also indicates that defense thresholds, above which herbivore performance declines precipitously, are uncommon, which implies that increasing plant investment in a given defense will consistently decrease herbivore performance.

Second, we asked how these relationships differed among herbivore species with different resource-selection behaviors. Because the consequences of plant variance for an herbivore population depend not only on the shape of the herbivore performance function but also on the amount of plant variance the population encounters, we hypothesized that curve shapes might be different for herbivore species with different mobility and host breadth—two traits that influence how herbivore species encounter plant variance. We found, however, that mobility did not influence the patterns described above for growth (nutrients: $\chi^2_1 = 0.26$, $P = 0.61$; defenses: $\chi^2_1 = 0.10$, $P = 0.75$) and survival (defenses: $\chi^2_2 = 0.45$, $P = 0.80$) (Extended Data Fig. 3). Herbivore host breadth also was not a significant predictor of Jensen's effect for growth (nutrients: $\chi^2_1 = 1.04$, $P = 0.31$; defenses: $\chi^2_1 = 0.18$, $P = 0.67$) or survival (defenses: $\chi^2_1 = 1.25$, $P = 0.26$) (Extended Data Fig. 4). Sample sizes were too small to test nutrient effects on survival. These results suggest that the shapes of nutrient and defense performance curves are fundamental constraints regardless of insect life history traits.

Our findings indicate that plants may contribute to the suppression of herbivore populations not only by having low average quality but also by having heterogeneity in nutrient levels. Concave-down nutrient-performance functions may be an important link between herbivore physiology and the negative relationships commonly observed between plant diversity and herbivore density at ecosystem scales²⁰. A key implication is that agroecosystems may experience outbreaks of herbivores because herbivore performance is elevated by artificially

low plant heterogeneity, due to landscape simplification, reduced plant species diversity, and crops bred to minimize variation. Increasing heterogeneity in plant nutrients in agroecosystems may be a key step towards the sustainable control of insect pests. Plant nutrient heterogeneity could be increased by planting greater numbers of crop varieties²¹, by increasing genetic diversity within crop varieties, or by breeding varieties with elevated constitutive or induced nutrient variance within plant parts attacked by insect pests.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 November 2015; accepted 5 October 2016.

Published online 12 October 2016.

1. Awmack, C. S. & Leather, S. R. Host plant quality and fecundity in herbivorous insects. *Annu. Rev. Entomol.* **47**, 817–844 (2002).
2. Adler, F. R. & Karban, R. Defended fortresses or moving targets? Another model of inducible defenses inspired by military metaphors. *Am. Nat.* **144**, 813–832 (1994).
3. Karban, R., Agrawal, A. A. & Mangel, M. The benefits of induced defenses against herbivores. *Ecology* **78**, 1351–1355 (1997).
4. Herrera, C. M. *Multiplicity in Unity: Plant Subindividual Variation & Interactions with Animals*. (University of Chicago Press, 2009).
5. Crutsinger, G. M. Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science* **313**, 966–968 (2006).
6. McArt, S. H. & Thaler, J. S. Plant genotypic diversity reduces the rate of consumer resource utilization. *Proc. R. Soc. B* **280**, 20130639 (2013).
7. Esquinas-Alcázar, J. Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat. Rev. Genet.* **6**, 946–953 (2005).
8. Díaz, S. & Cabido, M. Vive la difference: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* **16**, 646–655 (2001).
9. Wood, S. A. *et al.* Functional traits in agriculture: agrobiodiversity and ecosystem services. *Trends Ecol. Evol.* **30**, 531–539 (2015).
10. Whitham, T. G. in *Variable Plants and Herbivores in Natural and Managed Systems* (eds Denno, R. F. & McClure, M. S.) 15–41 (Academic Press, 1983).
11. Ruel, J. J. & Ayres, M. P. Jensen's inequality predicts effects of environmental variation. *Trends Ecol. Evol.* **14**, 361–366 (1999).
12. Bolnick, D. I. *et al.* Why intraspecific trait variation matters in community ecology. *Trends Ecol. Evol.* **26**, 183–192 (2011).
13. Jensen, J. L. W. V. Sur les fonctions convexes et les inegalites entre les valeurs moyennes. *Acta Math.* **30**, 175–193 (1906).
14. Raubenheimer, A. D., Lee, K. P. & Simpson, S. J. Does Bertrand's rule apply to macronutrients? *Proc. R. Soc. B* **272**, 2429–2434 (2005).
15. Simpson, S. J. & Raubenheimer, D. A multi-level analysis of feeding behaviour: the geometry of nutritional decisions. *Philos. Trans. R. Soc. B* **342**, 381–402 (1993).
16. Mertz, W. The essential trace elements. *Science* **213**, 1332–1338 (1981).
17. Bertrand, G. On the role of trace substances in agriculture. *Eighth Int. Congr. Appl. Chem.* **28**, 30–40 (1912).
18. Calabrese, E. J. & Baldwin, I. A. Toxicology rethinks its central belief. *Nature* **421**, 691–692 (2003).
19. Ali, J. G. & Agrawal, A. A. Specialist versus generalist insect herbivores and plant defense. *Trends Plant Sci.* **17**, 293–302 (2012).
20. Andow, D. A. Vegetational diversity and arthropod population response. *Annu. Rev. Entomol.* **36**, 561–586 (1991).
21. Zhu, Y. *et al.* Genetic diversity and disease control in rice. *Nature* **406**, 718–22 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Thaler, J. Rosenheim, A. Agrawal, and S. Ellner for comments on the manuscript. Our thinking benefited from discussion with D. Strong, K. Poveda, A. Kessler, S. Schreiber, P. Grof-Tisza, the University of California, Davis Center for Population Biology, and the Cornell University Department of Entomology. This work was supported by grants from the Center for Population at the University of California, Davis.

Author Contributions W.W. conceived the project. All authors contributed to the development of the question, interpreted the results, and commented on the manuscript. W.W., H.K., and M.R. collected data and assembled the database. W.W. and M.H. developed the methods. W.W. and R.K. wrote the manuscript. M.R., H.K., and W.W. made the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.W. (wcwetz@cornell.edu).

Reviewer Information *Nature* thanks M. Ayres, B. Inouye, W. Viechtbauer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

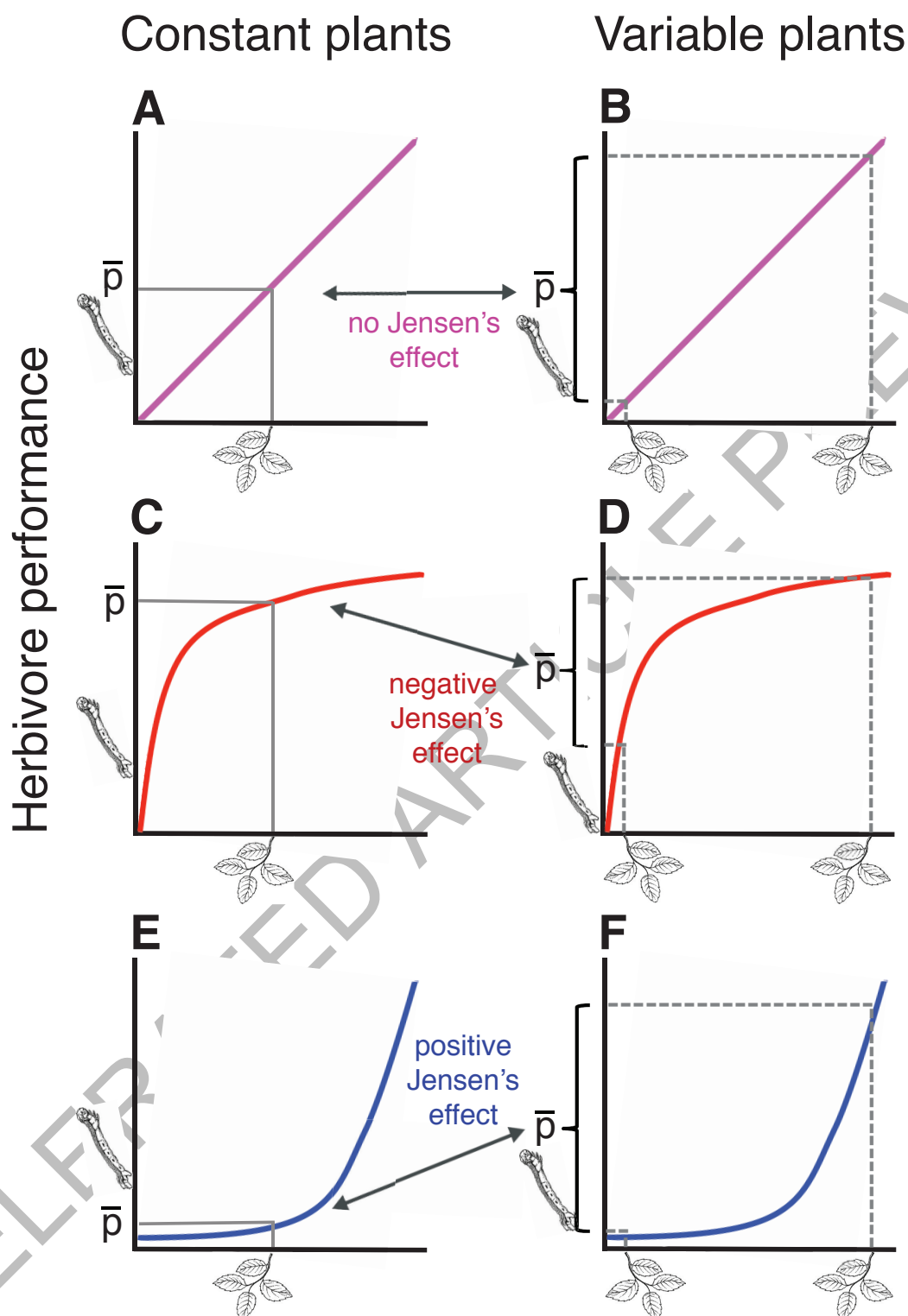


Figure 1 | An illustration of Jensen's inequality. The shape of the relationship between herbivore performance and a plant trait influences the consequences of trait variance for mean performance (\bar{p}) via nonlinear averaging or Jensen's inequality. The left column (A,C,E) represents plant populations where all plants have one trait value and no variance (plants at one location on x-axis). The right column (B,D,F) represents populations with trait variance, where half of the plants have a high trait value and half have a low value (two plants on x-axis). The trait mean, however, is the same in the constant and variable populations. With any linear function, trait value changes result in proportional changes in performance. Mean

herbivore performance is therefore equal in the absence or presence of variance (no Jensen's effect; compare A,B). With nonlinear performance functions (C-F), however, trait value changes do not result in proportional performance changes, and mean performance will differ in the absence and presence of trait variance. When the relationship is concave-down, mean performance will be lower in the presence of trait variance (negative Jensen's effect; compare C,D). When the relationship is concave-up, mean performance will be higher in the presence of trait variance (positive Jensen's effect; compare E,F). Doubled-headed arrows show differences in mean performance with and without plant variance.

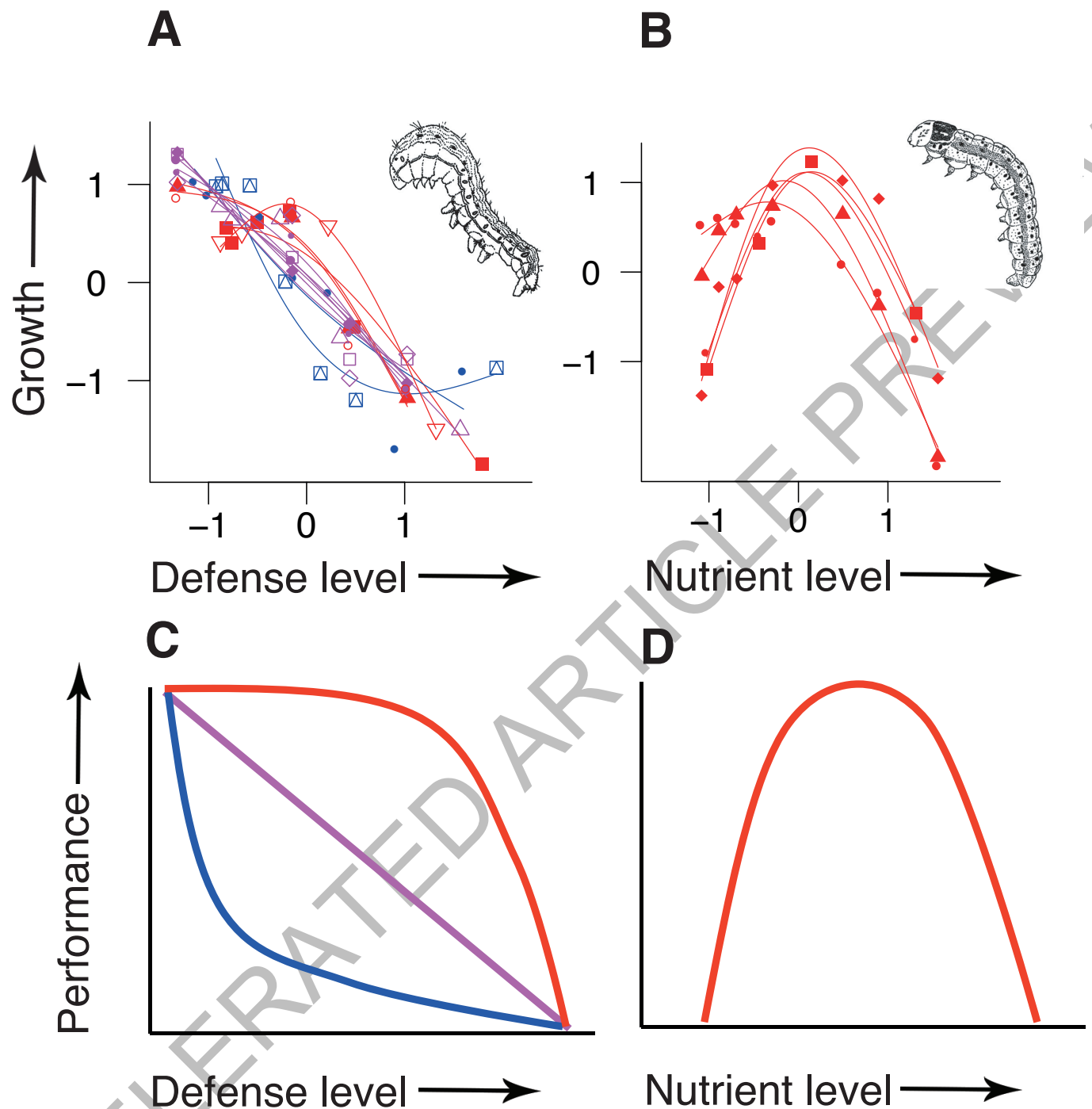


Figure 2 | Empirical and theoretical performance curves. Growth data from empirical studies and fitted growth curves for (A) *Heliothis virescens* (Lepidoptera: Noctuidae) across a range of levels of various plant secondary metabolites and (B) *Helicoverpa zea* (Lepidoptera: Noctuidae) across a range of levels of various plant nutrients. Data are standardized to zero mean and unit standard deviation. Red curves are significantly concave-down, blue ones are concave-up, and pink ones are

linear. Different symbols denote different experiments. (C,D) Theoretical predictions for relationships between plant traits and herbivore performance. For simplicity, (D) shows one curve with an intermediate maximum, but curves that asymptote at high nutrients are also possible. Both shapes are concave-down and would result in negative effects of nutrient variability.

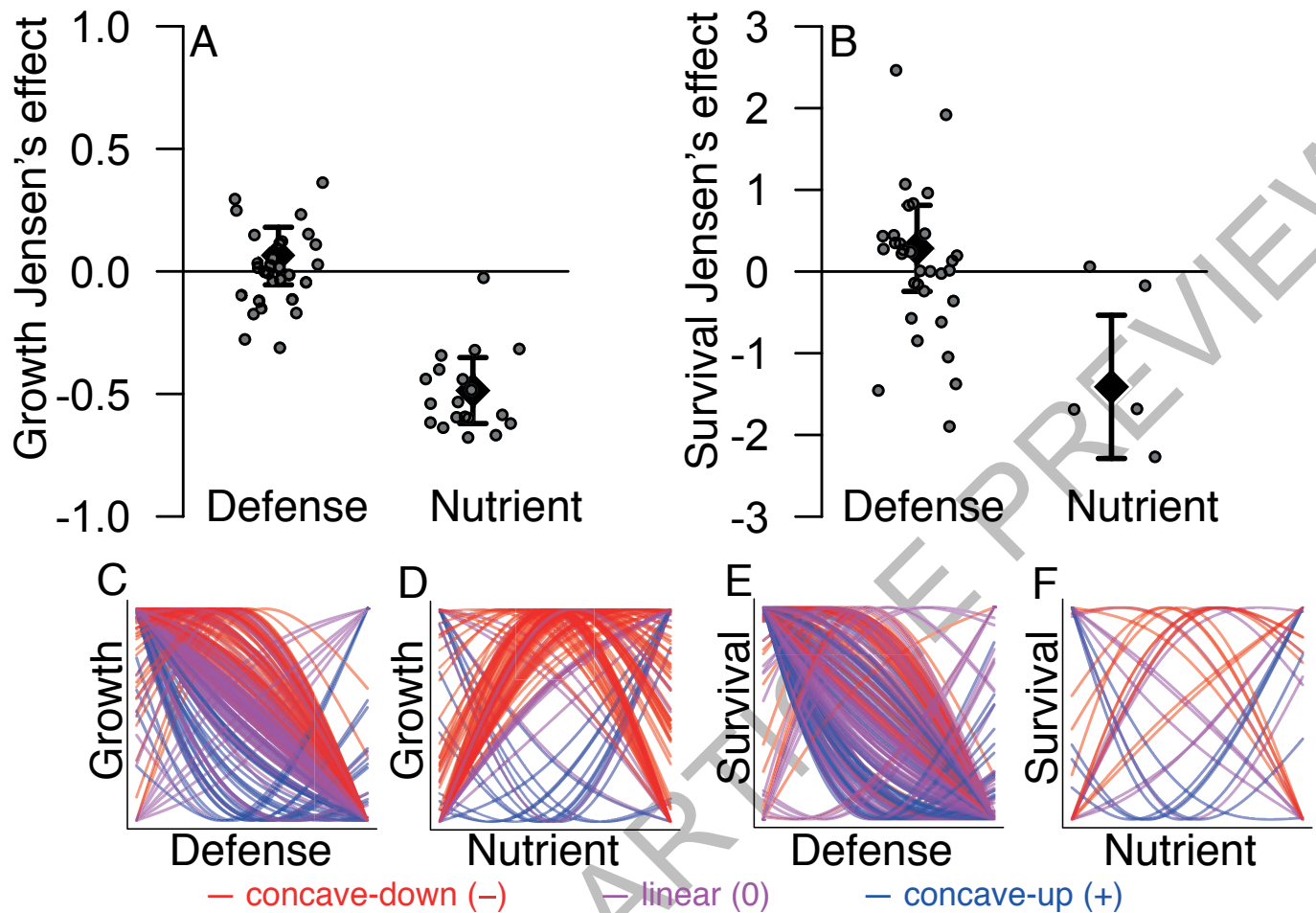


Figure 3 | The effect of variance in plant defensive and nutritive traits on herbivore growth (A) and survival (B). Each point is one herbivore species ($N = 53$), jittered for visibility. Diamonds and error bars show means and 95% confidence intervals. Growth effects are standard deviations. Survival effects are log odds-ratios. (C-F) The empirically-estimated

functions that went into the analysis that yielded (A) and (B). Red curves are significantly concave-down ($-$ Jensen's effect). Pink curves are linear (0 Jensen's effect). Blue curves are significantly concave-up ($+$ Jensen's effect). Curves are standardized to be on the same scale.

METHODS

Literature search. We located papers with data relating plant traits to herbivore performance by conducting keyword searches in Web of Science up to September 2014, collecting studies cited in relevant reviews,^{4,11,22} and searching papers known to the authors. Studies were included in our analysis if they met the following criteria: 1) a plant trait was experimentally manipulated and directly related to insect growth or survival; 2) at least four levels of the plant trait were established; 3) for all continuous response variables, some estimate of variance in herbivore performance was provided (e.g., SE) along with the mean and sample size at each treatment level; 4) for binomial survival response variables, studies provided an initial number of individuals at each treatment level and a count, proportion, or percentage that survived; and 5) data on traits and herbivore performance could be retrieved from a table, figure, text, or supplement²³. (See Supplementary Methods for additional methods.)

Data collection. From each suitable paper, we recorded species, plant traits, and herbivore growth and survival variables measured. We categorized plant traits as defenses or nutrients based on the original studies. (See Extended Data Fig. 1 for a summary of the database, including a list of all plant traits; see Supplementary Tables 1 and 2 for a list of the papers that met our criteria and a list of the herbivore species). For growth responses, we collected the mean and variability (e.g., SE) of the herbivore response and sample size at each level of the plant trait. For survival responses, we collected the initial number of herbivores and the count, proportion, or percent that survived at each level of the plant trait.

Effect calculation. We used a bootstrapping approach to calculate a distribution of Jensen's effects for each empirical observation (Extended Data Fig. 2). Having a distribution of effects for each observation allowed us to estimate variance for each observation and quantify our uncertainty for each estimate. For survival responses, we used nonparametric bootstrapping. We resampled each survival dataset with replacement 10,000 times. We fit cubic splines to each bootstrap dataset using the mgcv package in R 3.2.4^{24–26} and calculated a Jensen's effect from each spline. We calculated Jensen's effect as the log odds ratio of the mean of the predicted survival probabilities at each plant trait level and the predicted survival probability at the mean plant trait level. This measure, the log odds ratio, is widely used to express effect sizes in meta-analyses of response probabilities²⁷. The log odds ratio is beneficial because it puts survival on the logit scale, which accurately represents survival as a multiplicative process.

For growth data, which were typically reported as means and standard errors at each plant trait level instead of raw data, we used parametric bootstrapping (Extended Data Fig. 2). We parameterized a log-normal distribution for herbivore growth at each level of the plant trait using the reported herbivore performance means and standard errors. We then drew values from each distribution until the length of our bootstrap response vectors equaled the sample sizes reported at each level of the plant trait. We repeated this procedure to obtain 10,000 bootstrap datasets. Then we followed the curve fitting methods described above but first log-transformed the responses and then used a Gaussian error distribution. We calculated Jensen's effect for growth by subtracting the predicted herbivore performance for the mean level of the plant trait (the expected herbivore performance in the absence of plant trait variance) from the mean of the predicted herbivore performances at each plant trait level (the expected herbivore performance accounting for trait variance and nonlinear averaging). We standardized this difference by dividing it by the standard deviation of the bootstrapped herbivore performances to enable comparison across studies. This measure thus expresses the effect of nonlinear averaging in terms of standard deviations of herbivore performance and is analogous to Hedges' *d*, one of the most widely used meta-analysis effect sizes²⁷. (See Supplementary Discussion: R Script for the computer scripts used in this analysis.)

Effect of Trait Distribution. Our approach assumes that the values of the plant traits tested by the authors of each study reflect the natural distribution of trait values. We believe this is justified because most authors stated that they chose trait

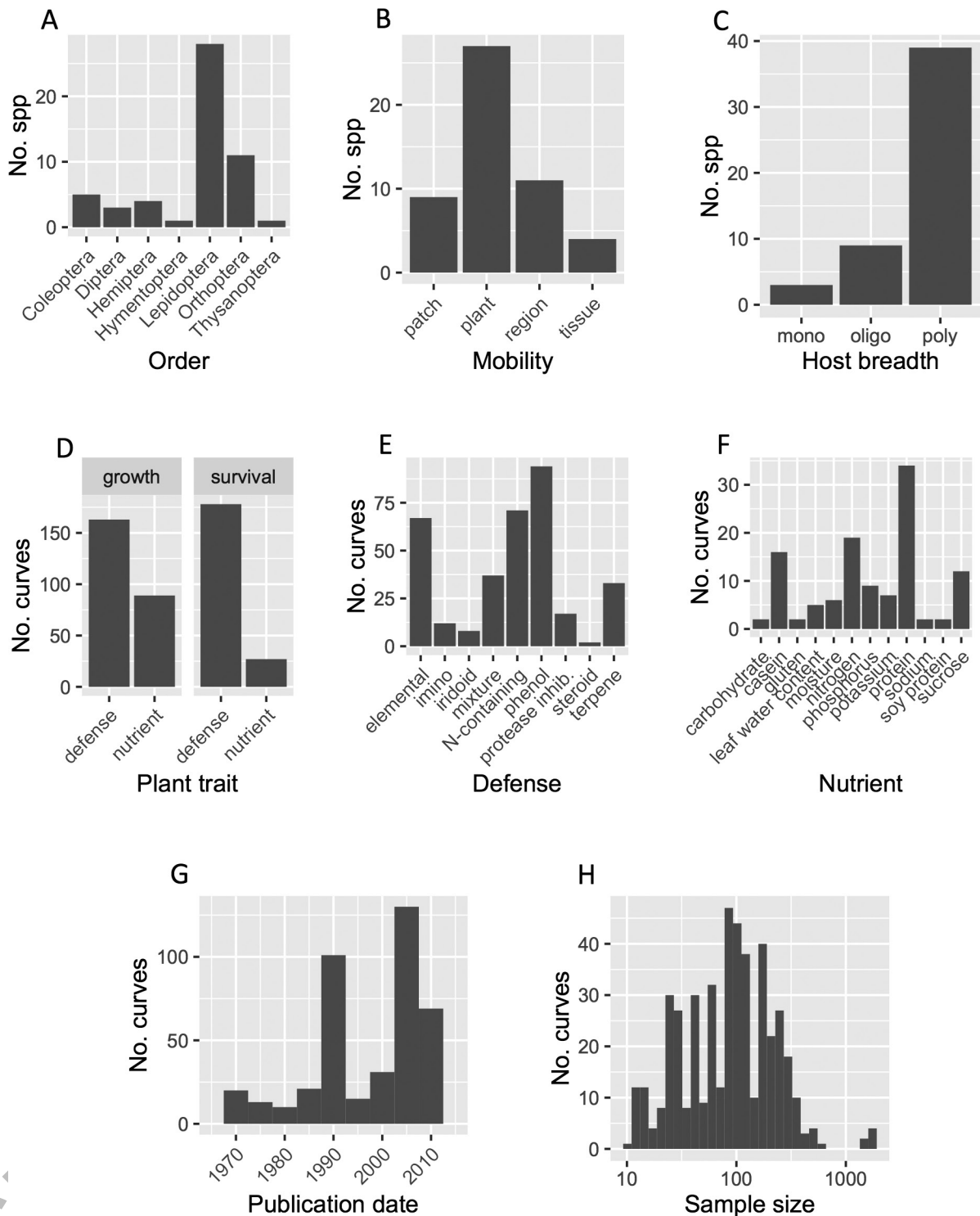
levels representative of those in nature or provided data showing correspondence between natural trait means and variances and experimentally chosen trait levels. Regardless, we also repeated the entire analysis assuming two different trait distributions: a uniform distribution between the minimum and maximum of the plant traits tested by the authors, and a Gaussian distribution with a mean equal to the midpoint of the author-chosen doses and a standard deviation that aligned the maximum author-chosen dose with 0.975 of the Gaussian cumulative probability function. The results were similar for each of the three trait distributions we tested, so in the main text we present the results based on the analysis that assumes a trait distribution defined by author-chosen trait levels. (See Supplementary Discussion for the results of the analyses with uniform and Gaussian distributions).

Statistical modeling. We tested our hypotheses using linear mixed effects models in the metafor package²⁸ in R 3.2.4²⁹. The response variable was the mean of the Jensen's effect distribution from each empirical dataset. We examined differences between variability in plant defensive and nutritive traits by fitting models with plant trait type (nutrient or defense) as an independent variable. We examined the effects of herbivore mobility and host-breadth by including mobility and host-breadth as independent variables. We tested the significance of trait type, mobility, and host-breadth as predictors of Jensen's effects using an omnibus test based on a chi-squared distribution. We used random intercepts for herbivore family and genus to account for potential correlations due to shared evolutionary history among genera within a family and species within a genus. We used this approach because a reliable phylogenetic tree does not exist for this diverse group of insects. We used an additional random intercept to account for the non-independence of multiple Jensen's effects measured on the same species, and we included a random effect for each observation following the standard practice of random effects meta-analysis³⁰. Finally, the sampling variance of each observation was set equal to the estimated variance of the distribution of Jensen's effects generated by our bootstrapping procedure. We explored differences in Jensen's effect among insect orders and found them to be minimal (Supplementary Discussion: Results by Insect Order).

Publication bias. Publication bias was unlikely to be an issue in our analysis because we used data for a different goal than did authors of the original studies, none of whom estimated Jensen's effects or curvature. Regardless, we explored the potential for publication bias graphically and found no evidence suggesting particular curve shapes were more likely to be published than others (Extended Data Figs. 5 and 6; Supplementary Discussion: Analysis of Publication Bias).

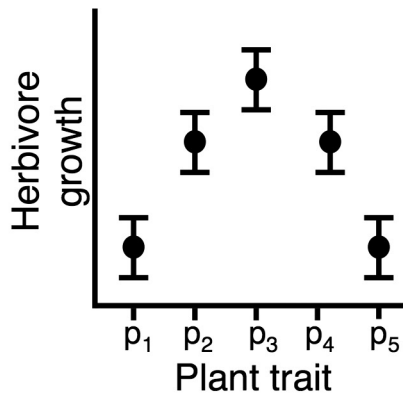
Data availability. Data used in the analysis have been deposited at <http://dx.doi.org/10.6084/m9.figshare.3792117>.

22. Carmona, D., Lajeunesse, M. J. & Johnson, M. T. J. Plant traits that predict resistance to herbivores. *Funct. Ecol.* **25**, 358–367 (2010).
23. Curtis, P. S., Mengersen, K., Lajeunesse, M. J., Rothstein, H. R. & Stewart, G. B. in *Handbook of Meta-analysis in Ecology and Evolution* (eds. Koricheva, J., Gurevitch, J. & Mengersen, K.) (Princeton University Press, 2013).
24. Wood, S. N. *Generalized Additive Models: An Introduction with R*. (Chapman and Hall/CRC, 2006).
25. R-Core-Team. R: A Language and Environment for Statistical Computing. (2015).
26. Schluter, D. Estimating the form of natural selection on a quantitative trait. *Evolution* **42**, 849–861 (1988).
27. Rosenberg, M. S., Rothstein, H. R. & Gurevitch, J. in *Handbook of Meta-analysis in Ecology and Evolution* (eds. Koricheva, J., Gurevitch, J. & Mengersen, K.) (Princeton University Press, 2013).
28. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
29. Bates, D., Maechler, M., Bolker, B. M. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. (2014).
30. *Handbook of Meta-analysis in Ecology and Evolution*. (Princeton University Press, 2013).

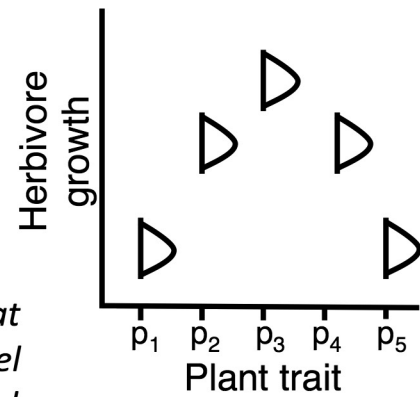


Extended Data Figure 1 | Graphical summary of database. Number of herbivore species per (A) order, (B) mobility of feeding stage, and (C) host breadth. Number of herbivore performance curves per (D) trait type, (E) defense class, (F) nutrient class, (G) date of publication, and (H) study sample size.

One per empirical data set

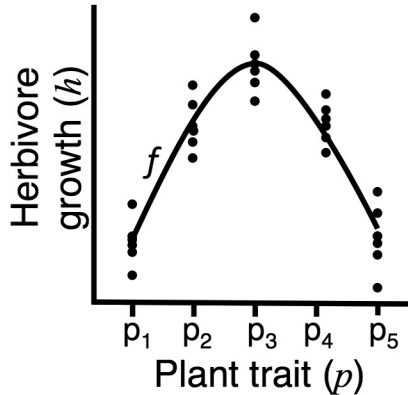


Parameterize distributions at each dose level using observed means and SEs

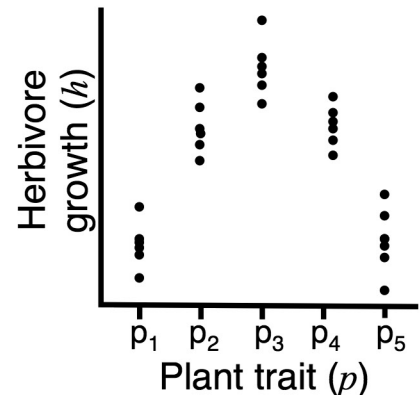


Draw n_i points from each distribution, where n_i is number of herbivores tested at dose i in original study

10,000 times per empirical data set



Approximate performance function using cubic spline f



Use f to predict mean herbivore performance on constant and variable plant populations

$$\bar{h}_c = f(\bar{p})$$

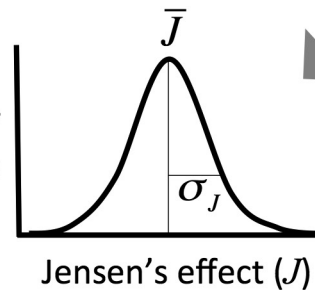
$$\bar{h}_v = \overline{f(p)}$$

Standardized Jensen's effect

$$J = \frac{\bar{h}_c - \bar{h}_v}{\sigma_h}$$

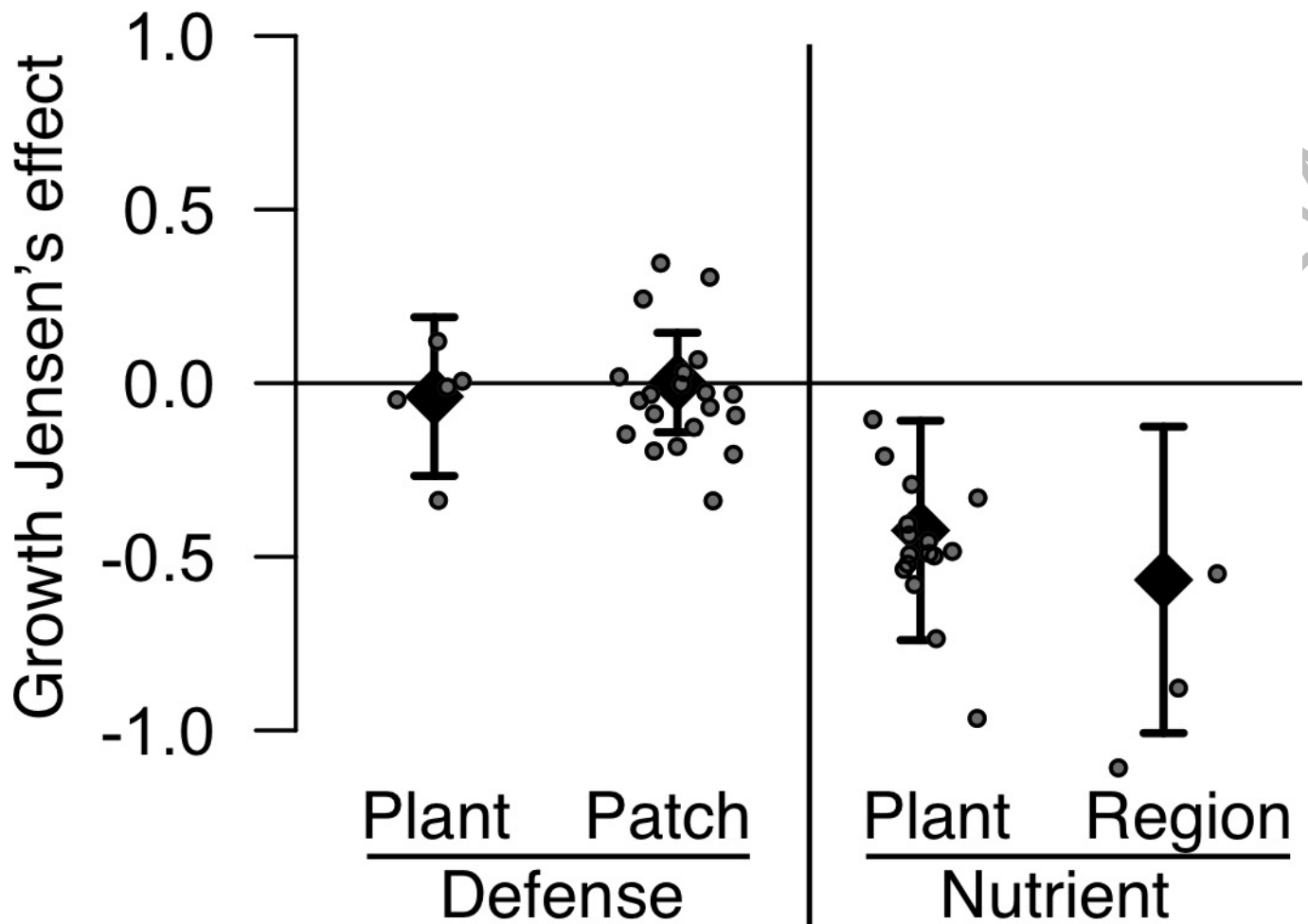
One per empirical data set

Number of bootstraps



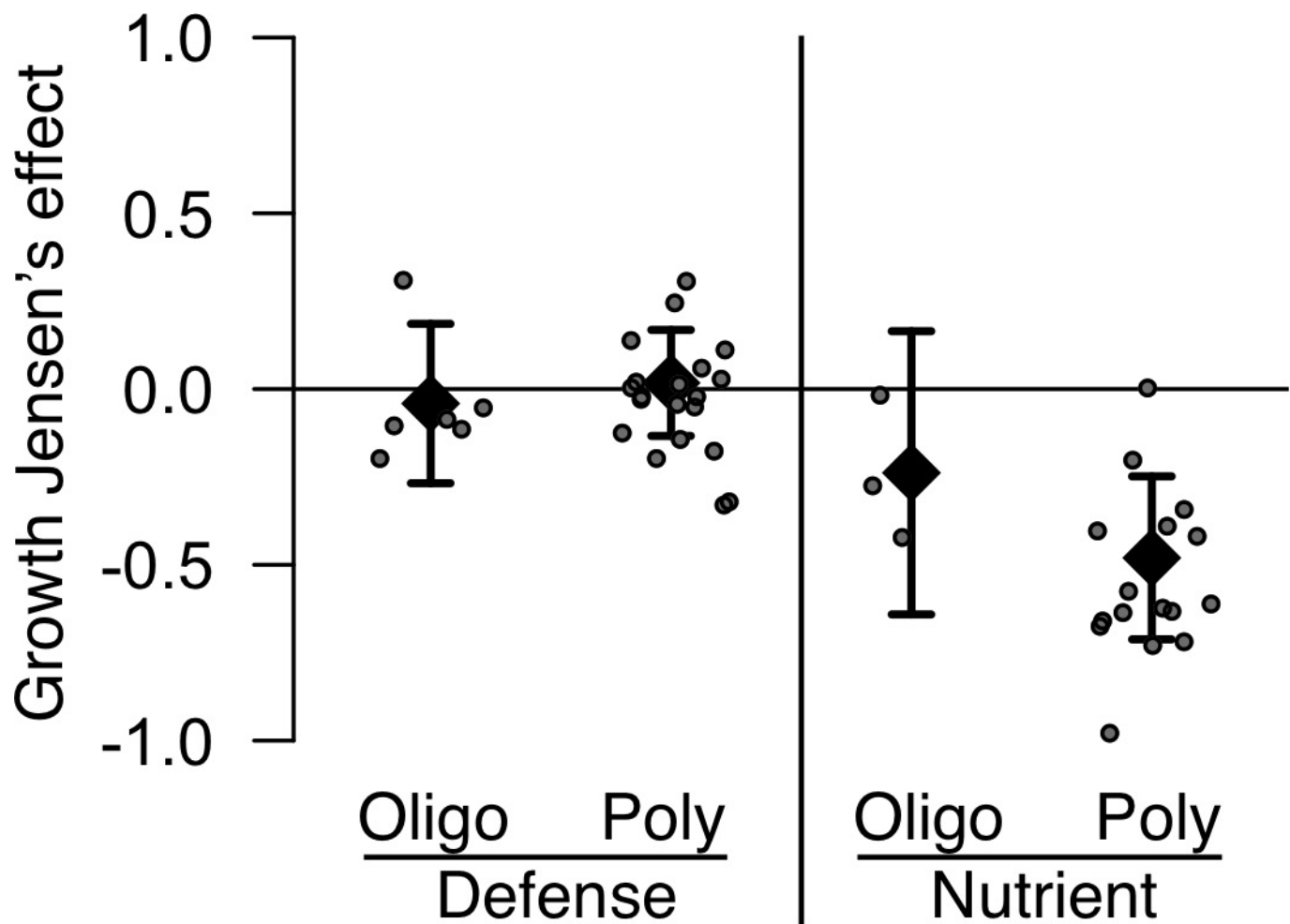
$\times 10,000$

Extended Data Figure 2 | Visual representation of quantitative methods. This diagram summarizes the bootstrapping algorithm used to calculate a distribution of Jensen's effects for each empirical data set for herbivore growth. For more details and for differences in methods between growth and survival see Methods and Supplementary Methods.



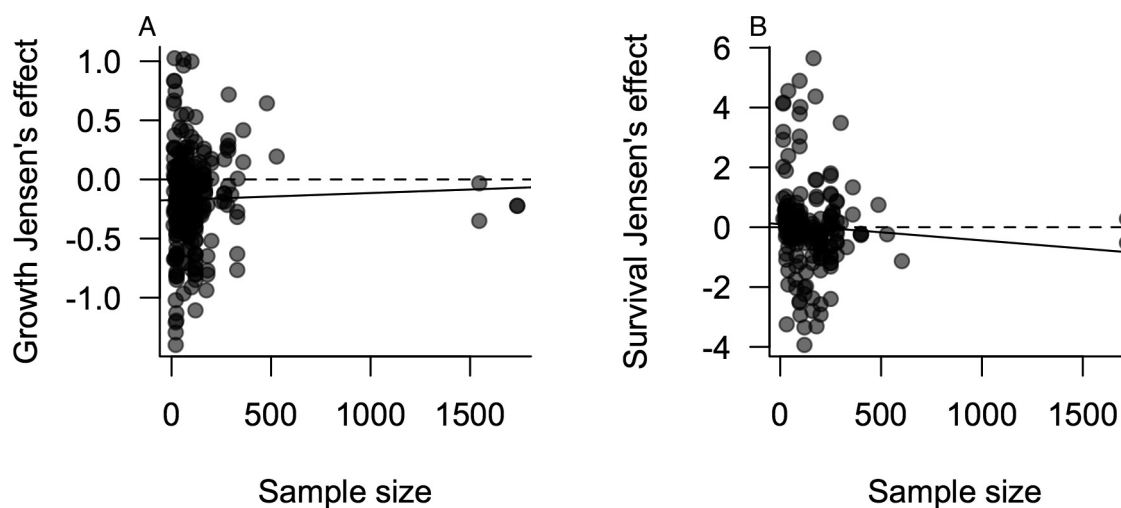
Extended Data Figure 3 | Jensen's effects by plant trait type (defenses and nutrients) and mobility of the feeding stage. Defense variance had mean effects near zero and nutrient variability had generally negative effects regardless of the mobility of the feeding stage of the herbivore species. Species in the 'plant' category move within plant individuals but do not typically move between plants. Species within the 'patch' category

readily move among neighboring host plants but do not typically move between patches. Species within the 'region' category commonly move among host plant patches. Each point is one herbivore species, jittered for visibility. Diamonds and error bars show means and 95% confidence intervals. See Supplementary Materials and Methods for more detail.

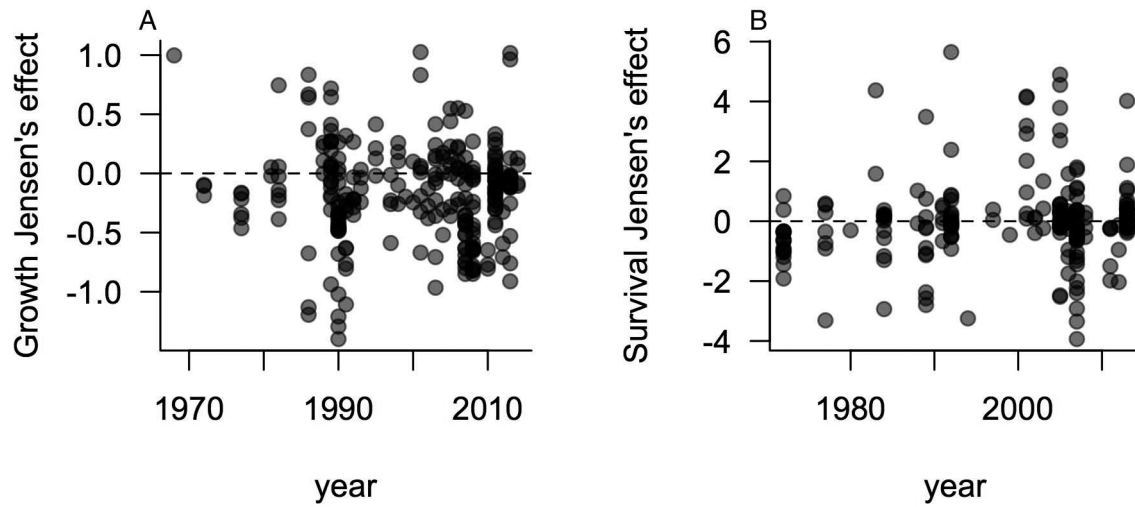


Extended Data Figure 4 | Jensen's effects by plant trait type (defenses and nutrients) and host-breadth. Defense variance had mean effects near zero and nutrient variability had generally negative effects regardless of the host breadth of the herbivore species. Oligophagous species ('oligo') feed on plant species in multiple genera but are restricted to one plant family.

Polyphagous species ('poly') feed on plant species across two or more plant families. Each point is one herbivore species, jittered for visibility. Diamonds and error bars show means and 95% confidence intervals. See Supplementary Materials and Methods for more detail.



Extended Data Figure 5 | Funnel plots for (A) growth and (B) survival. The lack of a relationship between the sample size of a study and its Jensen's effect suggests publication bias did not have a major influence on the results. Dashed line shows zero. Solid lines show linear regressions for growth ($F_{1,248} = 0.23$, $P = 0.63$, $R^2 = 0.0$) and survival ($F_{1,203} = 1.04$, $P = 0.31$, $R^2 = 0.0$).



Extended Data Figure 6 | Jensen's effect for each observation by the year of publication for (A) growth and (B) survival. The lack of temporal trends in Jensen's effects suggests publication bias did not play a major role.

Neuromodulators signal through astrocytes to alter neural circuit activity and behaviour

Zhiguo Ma¹, Tobias Stork¹, Dwight E. Bergles² & Marc R. Freeman^{1†}

Astrocytes associate with synapses throughout the brain and express receptors for neurotransmitters that can increase intracellular calcium (Ca^{2+})^{1–3}. Astrocytic Ca^{2+} signalling has been proposed to modulate neural circuit activity⁴, but the pathways that regulate these events are poorly defined and *in vivo* evidence linking changes in astrocyte Ca^{2+} levels to alterations in neurotransmission or behaviour is limited. Here we show that *Drosophila* astrocytes exhibit activity-regulated Ca^{2+} signalling *in vivo*. Tyramine and octopamine released from neurons expressing tyrosine decarboxylase 2 (Tdc2) signal directly to astrocytes to stimulate Ca^{2+} increases through the octopamine/tyramine receptor (Oct-TyrR) and the transient receptor potential (TRP) channel Water witch (Wtrw), and astrocytes in turn modulate downstream dopaminergic neurons. Application of tyramine or octopamine to live preparations silenced dopaminergic neurons and this inhibition required astrocytic Oct-TyrR and Wtrw. Increasing astrocyte Ca^{2+} signalling was sufficient to silence dopaminergic neuron activity, which was mediated by astrocyte endocytic function and adenosine receptors. Selective disruption of Oct-TyrR or Wtrw expression in astrocytes blocked astrocytic Ca^{2+} signalling and profoundly altered olfactory-driven chemotaxis and touch-induced startle responses. Our work identifies Oct-TyrR and Wtrw as key components of the astrocytic Ca^{2+} signalling machinery, provides direct evidence that octopamine- and tyramine-based neuromodulation can be mediated by astrocytes, and demonstrates that astrocytes are essential for multiple sensory-driven behaviours in *Drosophila*.

Astrocytes are intimately associated with brain synapses and positioned to broadly regulate synaptic activity. It has been widely proposed that astrocytes modulate neural circuits and behaviour^{5–7}, although *in vivo* evidence that astrocytes are directly activated by neurotransmission and signal back to neurons to modulate their output remains elusive. Astrocytes exhibit dynamic fluctuations in intracellular Ca^{2+} *in vitro*^{8,9} and *in vivo*^{10,11}, suggesting that Ca^{2+} signalling might be a useful measure of astrocytic activity. Despite decades of studies on astrocytic Ca^{2+} transients, the signalling pathways that control these transients remain poorly defined, and their *in vivo* relevance remains controversial. If astrocytes actively participate in information processing in circuits, it is imperative that we characterize the mechanisms involved as they would represent a potentially widespread mechanism for controlling brain function.

We performed an RNA interference (RNAi)-based screen in *Drosophila* in which we individually knocked down around 500 Ca^{2+} signalling-related genes selectively in astrocytes using the astrocyte-specific *Gal4* driver line *alm-Gal4* (ref. 12), and assayed larval olfactory-driven chemotaxis (Extended Data Fig. 1a). We found that astrocyte knockdown of the TRP channel Wtrw¹³ led to a significant (~50%; $P < 0.01$) decrease in larval chemotaxis towards iso-amyl acetate (IAA) (Fig. 1a). Similar results were found with a second, non-overlapping RNAi construct (Fig. 1a); the effect of Wtrw knockdown was specific

to glia, as shown by insensitivity to blockade of *Gal4/UAS* in neurons expressing *elav-Gal80* (Extended Data Fig. 1e). Larvae bearing the null allele *wtrw^{ex}* were equally defective in chemotaxis, but this behaviour was rescued by re-expression of *wtrw* only in astrocytes (Fig. 1a). We also found that blockade of *Gal4/UAS*-driven *wtrw^{RNAi}* with *tsh-Gal80* in the ventral nerve cord (VNC) resulted in normal chemotaxis responses (Extended Data Fig. 1d, f), revealing a critical role for VNC astrocytes in this behaviour, although this does not exclude an additional role for brain astrocytes. Chemotaxis defects did not result from simple alterations in motility, as *wtrw^{RNAi}* larvae exhibited normal locomotion (Extended Data Fig. 1g) and light avoidance responses (Extended Data Fig. 1c, h). We conclude that Wtrw is required in astrocytes for normal larval olfactory-driven behaviour.

Ca^{2+} transients are induced in astrocytes in awake behaving mice during periods of elevated arousal^{14–16}. We therefore used a gentle anterior touch assay¹⁷ to investigate the role of astrocyte Ca^{2+} signalling in larval startle-induced behaviours. Crawling larvae touched anteriorly with a hair responded by pausing and continuing forwards (type I response), or by moving backwards and executing an escape response (type II response; Extended Data Fig. 1b). Wild-type larvae and controls without RNA knockdown exhibited roughly equal frequencies of type I and type II responses, but the expression of *wtrw^{RNAi}* in astrocytes resulted in a marked alteration in behaviour: about 80% of larvae exhibited type I responses. This phenotype was mimicked by *wtrw^{ex}* mutants (Fig. 1b) and was independent of neuronal Wtrw expression (Extended Data Fig. 1i). These data indicate that astrocyte-expressed Wtrw also modulates startle-induced behavioural changes in *Drosophila* larvae.

To study *in vivo* Ca^{2+} signalling, we developed a semi-dissected preparation to image the larval CNS. We used the fluorescent calcium sensor GCaMP6s to image astrocyte cytosolic Ca^{2+} changes (UAS-GCaMP6s) and mCherry (UAS-mCherry) as a reference for astrocyte position, and imaged dorsal astrocyte cell bodies in the VNC. Astrocyte cell bodies exhibited coordinated, population-wide slow oscillations (termed somatic Ca^{2+} transients; Extended Data Fig. 1j–l and Supplementary Video 1). Somatic Ca^{2+} transients occurred approximately every 2 min and exhibited an average change in $\Delta F/F$ of about 11% (Extended Data Fig. 1m). Notably, blocking neuronal activity with tetrodotoxin (TTX) suppressed transients by approximately 60%, as did application of the broad Ca^{2+} channel blocker lanthanum chloride (LaCl_3 ; Extended Data Fig. 1n). Similar astrocytic Ca^{2+} transients were observed when we imaged intact immobilized larvae (Extended Data Fig. 1o), indicating that our dissected preparation preserves *in vivo* patterns of astrocyte activity.

TRP channels regulate Ca^{2+} levels in astrocytes¹⁸, so we reasoned that Wtrw might drive Ca^{2+} signalling in *Drosophila* astrocytes. Wild-type larvae exhibited 8–9 rhythmic oscillations in somatic Ca^{2+} transients over 15 min. By contrast, astrocyte-specific *wtrw^{RNAi}* led to a roughly 50% decrease in somatic astrocyte Ca^{2+} transients, which was also observed in the *wtrw^{ex}* mutant (Fig. 1c). Bath application of

¹Department of Neurobiology and Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ²Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [†]Present address: Vollum Institute, Oregon Health and Sciences University, Portland, Oregon 97239, USA.

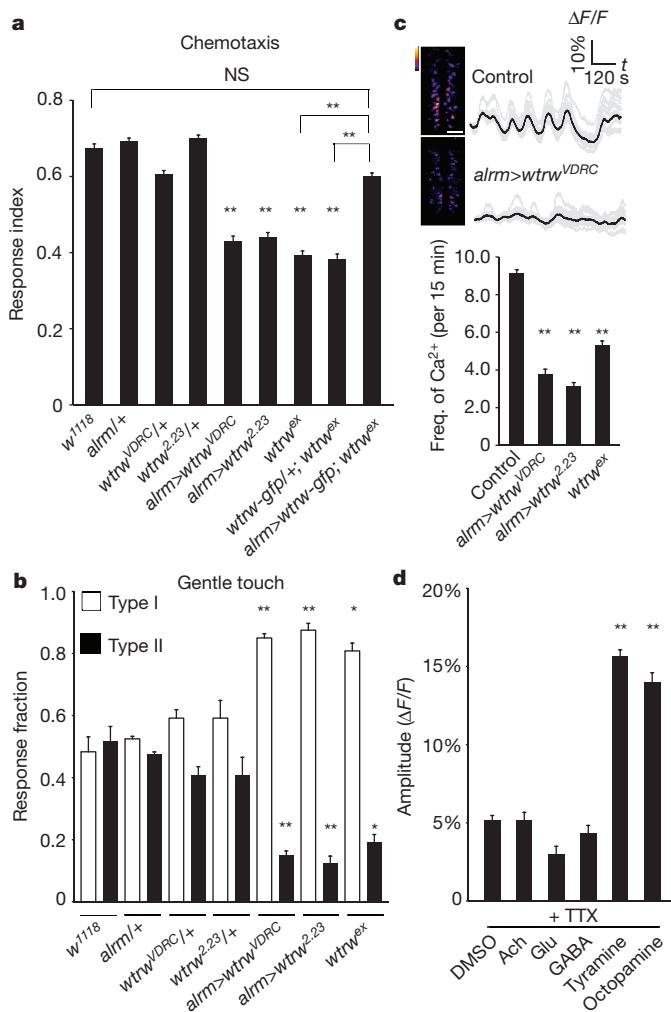


Figure 1 | Larval chemotaxis and startle-induced responses require the astrocyte-expressed TRP channel Wtrw. **a**, Chemotaxis assay ($n = 12$). **b**, Gentle touch assay ($n = 30$). **c**, Pseudocoloured maximum intensity projections of 15-min movies, averaged traces of 16 individual astrocytes and quantifications of the frequency of somatic Ca^{2+} transients ($n = 10$, 160 cells total). Scale bar, 50 μm . **d**, Responses of astrocytes to neurotransmitters and neuromodulators in the presence of TTX ($n = 6$, 96 cells total). * $P < 0.05$, ** $P < 0.01$; NS, not significant; error bars, s.e.m. **a**, Wilcoxon and Mann–Whitney tests followed by Bonferroni–Holm *post hoc* test; **b–d**, One-way ANOVA followed by Tukey’s *post hoc* test. Ach, acetylcholine; GABA, γ -aminobutyric acid; Glu, glutamate.

acetylcholine, glutamate, or γ -aminobutyric acid in the presence of TTX did not elicit a change in Ca^{2+} levels in astrocytes. By contrast, application of tyramine (Tyr) or octopamine (Oct), the invertebrate analogues of norepinephrine, which has been shown to induce Ca^{2+} transients in mammalian astrocytes^{14,15,19,20}, potentially elevated somatic Ca^{2+} in astrocytes (Fig. 1d), indicating that astrocyte somatic Ca^{2+} signalling is regulated by these neuromodulators.

Tdc2⁺ neurons are the only known source of Tyr and Oct in the larval VNC²¹. To investigate the relationship of these neurons with astrocytes we expressed the red-shifted Ca^{2+} indicator R-GECO1 in Tdc2⁺ neurons (using *tdc2-LexA/LexAop-R-GECO1*) and GCaMP6s in astrocytes (using *alrm-Gal4/UAS-GCaMP6s*) and examined *in vivo* activity. We observed a striking positive correlation between Tdc2⁺ neuron activity and somatic astrocyte Ca^{2+} : when Tdc2⁺ neurons were active, astrocyte somatic Ca^{2+} levels increased, and when Tdc2⁺ neurons were silent, astrocyte somatic Ca^{2+} levels decreased (Fig. 2a and Supplementary Video 2). A similar correlation was observed in intact larvae (Extended Data Fig. 1p). Moreover, the amplitude and duration of the increase in

somatic astrocyte Ca^{2+} was tightly correlated with Ca^{2+} spikes in Tdc2⁺ neurons (Fig. 2a). When we chronically silenced Tdc2⁺ neurons by expressing the K⁺ leak channel Kir2.1 (ref. 22), rhythmic oscillations in astrocyte Ca^{2+} were eliminated (Fig. 2b and Extended Data Fig. 2a); in addition, acute optogenetic blockade of Tdc2⁺ neuron activity using halorhodopsin (*tdc2-Gal4 > UAS-NpHR*) led to a decrease in astrocyte somatic Ca^{2+} (Fig. 2c and Extended Data Fig. 2b). Notably, astrocytic Ca^{2+} signalling was also blocked by the $\alpha 1$ adrenergic receptor antagonist terazosin (Fig. 2d and Extended Data Fig. 2c), which inhibits arousal-evoked Ca^{2+} responses in cortical astrocytes in mammals^{14,15}, further supporting the notion that astrocyte modulation is highly conserved between species. Consistent with astrocyte Ca^{2+} signalling being regulated by Oct and Tyr, somatic astrocyte Ca^{2+} transients were nearly absent (about 80% reduced) in *tdc2^{RO54}* mutants, which lack both Tyr and Oct, although they persisted in *t3h^{nM18}* mutants, which lack Oct but retain Tyr signalling (Fig. 2e, f). Finally, Tdc2⁺ neurons were activated when olfactory neurons were optogenetically stimulated in intact larvae (using *or83b-Gal4*), suggesting that sensory cues flow to Tdc2⁺ neurons, which are upstream of astrocyte Ca^{2+} signalling (Extended Data Fig. 2d).

To determine whether Tdc2⁺ neurons signal directly to astrocytes we knocked down the two Tyr receptors, TyrR and TyrRII, and Oct-TyrR, a dual specificity receptor that can bind both Oct and Tyr^{23,24}, in astrocytes and assayed Ca^{2+} signalling. Depletion of TyrR or TyrRII had no effect (data not shown), but depletion of Oct-TyrR strongly suppressed astrocyte somatic Ca^{2+} transients. We observed strong inhibition of astrocyte somatic Ca^{2+} transients in the *Oct-TyrR^{hono}* homozygous mutant, and in *Oct-TyrR^{hono}/+* heterozygous mutants, which were previously shown to have dominant effects²⁵ (Fig. 2g). Loss of one copy of *Oct-TyrR* (*Oct-TyrR^{hono}/+*) also led to defects in chemotaxis behaviour similar to those observed after astrocyte-specific knockdown of *wtrw*; these behavioural changes were enhanced in the *Oct-TyrR^{hono}* homozygous mutant, or by astrocyte-specific knockdown of *Oct-TyrR* with two copies of an *Oct-TyrR* RNAi construct. Co-expression of *Oct-TyrR^{RNAi}* and *wtrw^{RNAi}* did not enhance chemotaxis defects, suggesting that Oct-TyrR and Wtrw act in a common genetic pathway (Fig. 2h). Finally, astrocyte-specific depletion of Oct-TyrR led to strong inhibition of type II escape responses in the gentle anterior touch assay (Fig. 2i). We conclude that Wtrw and Oct-TyrR are required in astrocytes to trigger somatic Ca^{2+} transients in response to Oct and Tyr, and astrocytic Wtrw and Oct-TyrR are essential for normal olfactory-driven chemotaxis and startle-induced escape responses.

Astrocytes support neuronal function in many ways and it is plausible that alterations in astrocyte Ca^{2+} signalling might modify upstream Tdc2⁺ neuron firing to influence behaviour. We therefore measured the frequency and amplitude of firing events in Tdc2⁺ neurons in *wtrw^{ex}* null mutants and found no alterations compared to wild-type controls (Extended Data Fig. 2e). That Tdc2⁺ neurons fire normally in *wtrw* mutants indicates that blocking astrocytic Ca^{2+} signalling does not lead to a global disruption of neuronal activity in the larval CNS.

Biogenic amines often regulate dopaminergic neurons²⁶. Triple labelling revealed that astrocytes, Tdc2⁺ neurons and dopaminergic neurons are in close proximity in the larval CNS (Extended Data Fig. 2f). When we imaged dopaminergic neuron activity (using *th-Gal4/UAS-GCaMP6s*) we found that *wtrw^{ex}* mutants exhibited a significant increase (about 50–70%; $P < 0.01$ for t1, $P < 0.05$ for t2 and t3) in the frequency of dopaminergic Ca^{2+} transients in multiple segmental clusters of tyrosine hydroxylase (TH)⁺ cells. A similar phenotype was observed in *Oct-TyrR^{hono}/+* mutants (Fig. 3a), although there were no significant changes in the average amplitude of dopaminergic neuron responses (Extended Data Fig. 2g). Thus Wtrw and Oct-TyrR negatively regulate dopaminergic neuron firing. Consistent with this observation, bath application of Oct or Tyr, but not vehicle (DMSO), to the larval CNS suppressed dopaminergic neuron activity within seconds (Fig. 3b). Strikingly, depletion of the Oct-TyrR and Wtrw specifically from astrocytes using RNAi potentially suppressed Tyr-induced Ca^{2+}

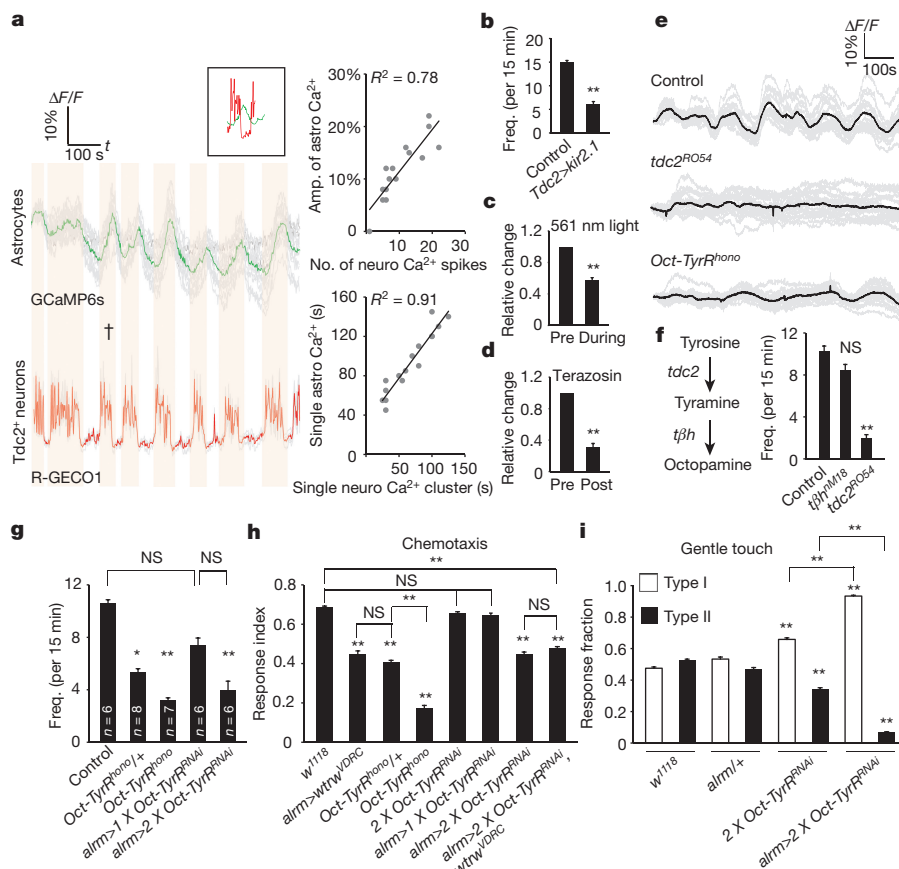


Figure 2 | Tdc2⁺ neurons fire rhythmically to drive astrocyte somatic Ca²⁺ transients through the Oct-TyrR. a, Correlation analyses of neuronal activity and somatic Ca²⁺ transients in astrocytes. Averaged traces of 16 individual astrocytes (GCaMP6s) and 4 pairs of Tdc2⁺ neurites (R-GEQ1) from the same sample. Vertical orange bars highlight concomitant activity in astrocytes and neurons (↑events shown in inset). Amplitude and duration of individual Ca²⁺ transients in astrocytes correlate highly with activity of Tdc2⁺ neurons. **b**, Frequency of Ca²⁺ transients in astrocytes (*n* = 10, 160 cells total). **c**, **d**, Relative changes in numbers of Ca²⁺ transients after acute blockade of Oct and Tyr signalling

elevation in astrocytes and silencing of dopaminergic neurons (Fig. 3c and Extended Data Fig. 2k). These data indicate that Tyr signals through astrocyte Oct-TyrR and Wtrw to inhibit downstream dopaminergic neuron activity.

To investigate how modulation of dopaminergic neurons regulates larval chemotaxis behaviour, we performed a series of acute manipulations. We first expressed the temperature-sensitive channel TrpA1 in dopaminergic neurons; after 24-h ectopic activation of dopaminergic neurons, larvae exhibited profound chemotaxis defects similar to those observed when *Wtrw* or Oct-TyrR were depleted from astrocytes (Extended Data Fig. 2h). In addition, feeding larvae a dopamine D1-like receptor antagonist rescued deficits associated with *wtrw*^{RNAi} in astrocytes in chemotaxis assays (Extended Data Fig. 2i). These data support the notion that the increased dopaminergic neuron activity observed after depletion of astrocytic Oct-TyrR and *Wtrw* is responsible for defects in chemotaxis behaviour.

How do astrocytes regulate dopaminergic neuron firing? Astrocytic release of ATP and signalling through purinergic receptors is one mechanism by which astrocytes appear to modulate neuronal activity^{27,28}. The *Drosophila* genome contains a single seven-transmembrane-domain purinergic receptor, AdoR, that is most similar to mammalian adenosine receptors²⁹. Pretreatment of larval CNS preparations with the AdoR antagonist SCH-442416 blocked Tyr-induced silencing of dopaminergic neurons without altering overall dopaminergic neuron

by halorhodopsin (**c**) or terazosin (**d**) ($n=6$, 96 cells total). Black dots denote points at which either 561 nm light or terazosin was administered. **e**, Representative traces of astrocytic Ca^{2+} transients in mutants defective in Oct and Tyr signalling. **f**, **g**, Frequency of Ca^{2+} transients in *tdc2^{RO54}*, *t3h^{nM18}* (**f**, $n=6$, 96 cells total) and *Oct-TyrR^{hono}* mutants (**g**, n listed for each genotype, 16 cells randomly selected and analysed for each sample). **h**, Chemotaxis assay ($n=12$). **i**, Gentle touch assay ($n=30$). * $P<0.05$, ** $P<0.01$; NS, not significant; error bars, s.e.m. **c**, **d**, Paired t -test; **b**, **f**–**i**, one-way ANOVA followed by Tukey's *post hoc* test.

activity before Tyr treatment. Similar results were observed in AdoR null mutants (Fig. 4a and Extended Data Fig. 2j). Purinergic signalling often results from vesicular release of ATP, which can be hydrolysed to adenosine by ectoenzymes. We manipulated exocytosis selectively in astrocytes using a dominant-negative version of the dynamin Shibire^{DN} (Shi^{DN}) and found that this suppressed the ability of Tyr to silence dopaminergic neurons. This blockade was not enhanced further by application of AdoR antagonists and had no effect on baseline dopaminergic neuron activity before Tyr treatment (Fig. 4b and Extended Data Fig. 2l). Finally, to determine whether astrocyte Ca²⁺ signalling through TRP channels is sufficient to suppress dopaminergic neuron firing, we expressed the wasabi (AITC)-sensitive *Drosophila* TrpA1 channel in astrocytes and tested the effects of AITC application on astrocytic Ca²⁺ signalling and dopaminergic neuron firing. Exposure of control preparations to AITC did not alter astrocyte Ca²⁺ levels, but application of AITC to preparations expressing TrpA1 in astrocytes led to an increase in astrocytic GCaMP6s signals of around 20% (Extended Data Fig. 2m), similar to that elicited by Tyr application (Fig. 1d). Moreover, activation of astrocyte Ca²⁺ signalling by AITC resulted in strong suppression of dopaminergic neuron activity. This effect was reversed by application of AdoR antagonists and did not occur when AITC was applied to control (that is, non-TrpA1 expressing) preparations (Fig. 4c and Extended Data Fig. 2n). Thus TrpA1 channel-mediated increases in astrocytic Ca²⁺ signalling are sufficient to silence

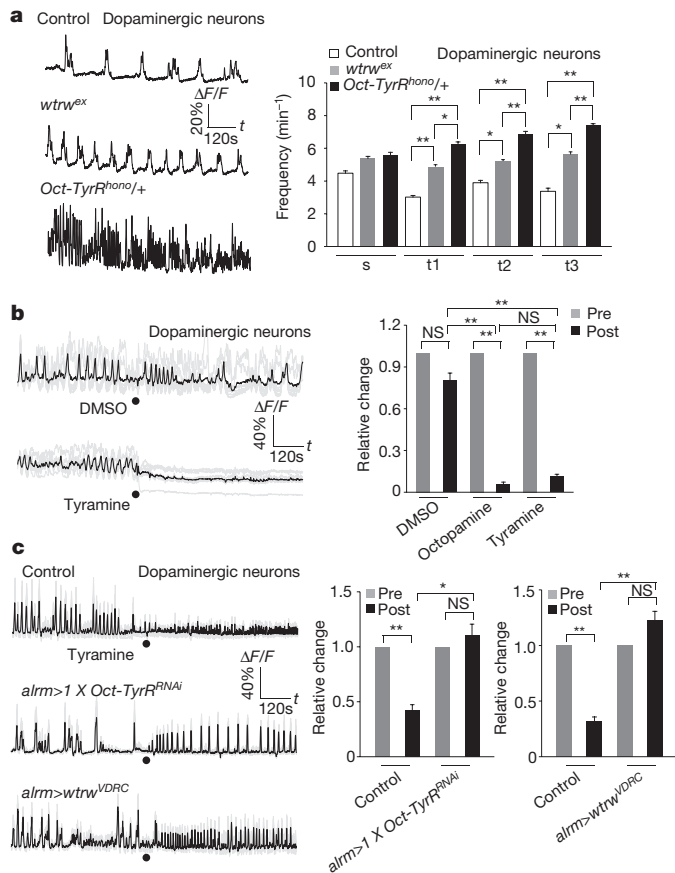


Figure 3 | Astrocytes mediate Tyr-induced inhibition of dopaminergic neurons through Oct-TyrR. **a**, Enhanced activity of dopaminergic neurons in *wtrw^{ex}* mutant and *Oct-TyrR^{homo/+}* heterozygote larvae ($n = 10$; 80 neurites). s, subesophageal segments; t, thoracic segments. **b**, Tyr and Oct (2.5 mM) inhibit the activity of dopaminergic neurons ($n = 6$; 48 neurites). **c**, Astrocyte-specific RNAi for *Oct-TyrR* or *wtrw* attenuates inhibition of dopaminergic neurons by Tyr ($n = 6$; 48 neurites). Black dots, points at which Tyr was perfused. * $P < 0.05$, ** $P < 0.01$; NS, not significant; error bars, s.e.m. **b**, **c**, Paired t -test (**b**, same treatment; **c**, same genotype); **a**–**c**, one-way ANOVA followed by Tukey's *post hoc* test.

dopaminergic neurons *in vivo*. Silencing of dopaminergic neurons is likely to be mediated by release of ATP from astrocytes and subsequent activation of AdoR on dopaminergic neurons.

It has been assumed that neuromodulators alter neural circuit activity and behaviour by acting directly on neurons; however, astrocytes (and other glia) express a diverse array of neuromodulatory receptors, suggesting they may contribute to the global effects of these neurotransmitters. In this study we have demonstrated that neuromodulatory signalling in at least some cases flows through astrocytes. We have shown that *Drosophila* astrocytes exhibit robust Ca^{2+} signalling events that are remarkably similar to those observed in awake behaving mice^{14–16}. Similar to the global activation of astrocyte Ca^{2+} signalling by norepinephrine in mammals, release of Tyr (and probably Oct) *in vivo* triggers synchronous activation of astrocytes. We have identified Wtrw and Oct-TyrR as key components of the astrocyte Ca^{2+} signalling axis, and their loss-of-function phenotypes have provided direct evidence that astrocytic Ca^{2+} signalling modulates animal behaviour. Furthermore, we have shown that, during suppression of dopaminergic neuron activity, Oct and Tyr signal directly to astrocytes, and astrocyte Ca^{2+} transients execute neuromodulatory events downstream (Fig. 4d). Oct and Tyr have been linked to arousal and aggression in insects and are considered to be functionally equivalent to norepinephrine in mammals. The profound modulation of astrocytic Ca^{2+} levels by norepinephrine in mammals^{14–16} suggests that activation of

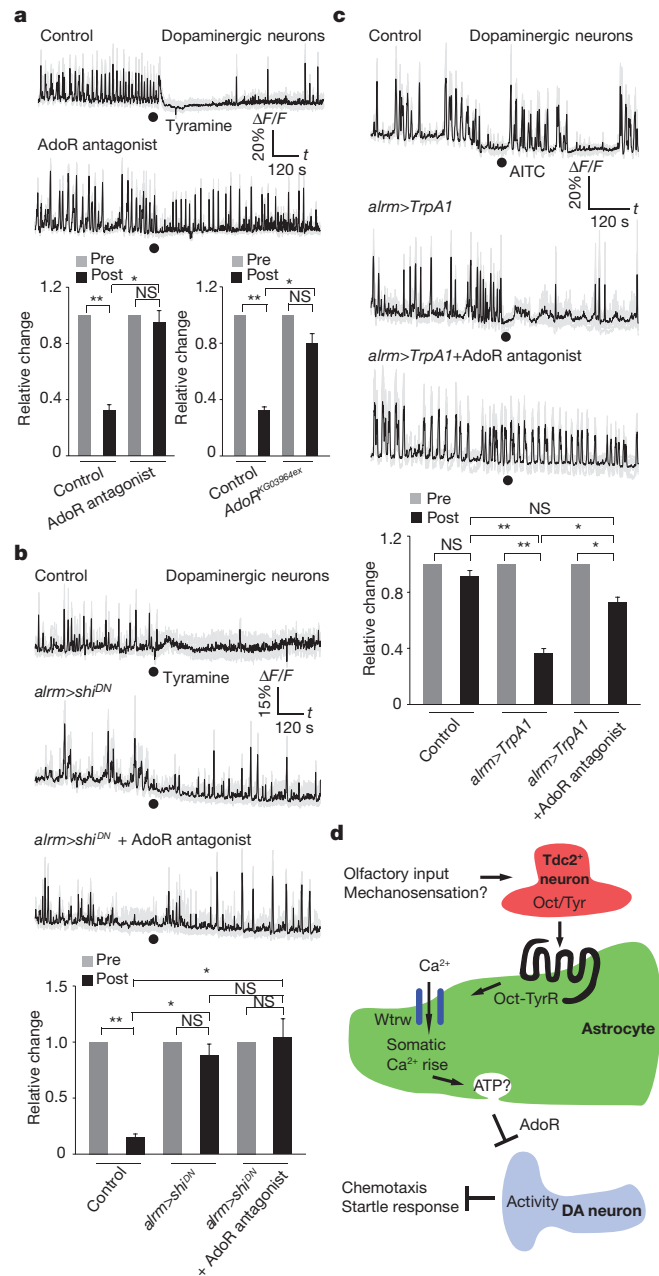


Figure 4 | Tyr-mediated inhibition of dopaminergic neurons depends on adenosine receptors and glial endocytic function. **a**, AdoR is required for Tyr-mediated inhibition of dopaminergic neurons ($n = 6$; 48 neurites). **b**, Blockade of endocytic function by dominant-negative shibire (*shib^{DN}*) attenuates inhibition of dopaminergic neurons by Tyr ($n = 6$; 48 neurites). Black dots, points at which Tyr was perfused. **c**, Ca^{2+} influx to astrocytes through TrpA1 is sufficient to inhibit dopaminergic neurons ($n = 6$; 48 neurites). Black dots, points at which AITC was perfused. **d**, Model. Olfactory and (probably) mechanosensory information flow towards Tdc2⁺ neurons, which release Oct and Tyr to activate Oct-TyrR on astrocytes and in turn induce astrocytic Ca^{2+} entry through the TRP channel Wtrw. The increase in astrocytic Ca^{2+} is sufficient to silence dopaminergic (DA) neurons through a mechanism requiring the adenosine receptor AdoR, potentially through astrocyte ATP release and its breakdown to adenosine. Inhibition of dopaminergic neuron activity by astrocytes is essential for normal chemotaxis and startle-induced reversal. * $P < 0.05$, ** $P < 0.01$; NS, not significant; error bars, s.e.m. **a**–**c**, Paired t -test (same genotype); Wilcoxon and Mann–Whitney tests followed by Bonferroni–Holm *post hoc* test (**a**, **b**); one-way ANOVA followed by Tukey's *post hoc* test (**c**).

astrocytic Ca^{2+} signalling by Tyr, Oct or norepinephrine, and in turn astrocyte-based neuromodulation, are ancient features of the metazoan nervous system. A reassessment of the cellular basis of neuromodulation is therefore warranted, with further reflection on direct roles for astrocytes in transducing neuromodulatory signals in neural circuits.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 October 2015; accepted 7 October 2016.

Published online 9 November 2016.

1. Cornell-Bell, A. H., Finkbeiner, S. M., Cooper, M. S. & Smith, S. J. Glutamate induces calcium waves in cultured astrocytes: long-range glial signaling. *Science* **247**, 470–473 (1990).
2. Charles, A. C., Merrill, J. E., Dirksen, E. R. & Sanderson, M. J. Inter cellular signaling in glial cells: calcium waves and oscillations in response to mechanical stimulation and glutamate. *Neuron* **6**, 983–992 (1991).
3. Dani, J. W., Chernjavsky, A. & Smith, S. J. Neuronal activity triggers calcium waves in hippocampal astrocyte networks. *Neuron* **8**, 429–440 (1992).
4. Smith, S. J. Do astrocytes process neural information? *Prog. Brain Res.* **94**, 119–136 (1992).
5. Khakh, B. S. & McCarthy, K. D. Astrocyte calcium signaling: from observations to functions and the challenges therein. *Cold Spring Harb. Perspect. Biol.* **7**, a020404 (2015).
6. Araque, A. *et al.* Gliotransmitters travel in time and space. *Neuron* **81**, 728–739 (2014).
7. Khakh, B. S. & Sofroniew, M. V. Diversity of astrocyte functions and phenotypes in neural circuits. *Nat. Neurosci.* **18**, 942–952 (2015).
8. Fatatis, A. & Russell, J. T. Spontaneous changes in intracellular calcium concentration in type I astrocytes from rat cerebral cortex in primary culture. *Glia* **5**, 95–104 (1992).
9. Nett, W. J., Oloff, S. H. & McCarthy, K. D. Hippocampal astrocytes in situ exhibit calcium oscillations that occur independent of neuronal activity. *J. Neurophysiol.* **87**, 528–537 (2002).
10. Porter, J. T. & McCarthy, K. D. Hippocampal astrocytes in situ respond to glutamate released from synaptic terminals. *J. Neurosci.* **16**, 5073–5081 (1996).
11. Nimmerjahn, A., Mukamel, E. A. & Schnitzer, M. J. Motor behavior activates Bergmann glial networks. *Neuron* **62**, 400–412 (2009).
12. Doherty, J., Logan, M. A., Taşdemir, O. E. & Freeman, M. R. Ensheathing glia function as phagocytes in the adult *Drosophila* brain. *J. Neurosci.* **29**, 4768–4781 (2009).
13. Liu, L. *et al.* *Drosophila* hygro-sensation requires the TRP channels water witch and nanchung. *Nature* **450**, 294–298 (2007).
14. Ding, F. *et al.* α 1-Adrenergic receptors mediate coordinated Ca^{2+} signaling of cortical astrocytes in awake, behaving mice. *Cell Calcium* **54**, 387–394 (2013).
15. Paukert, M. *et al.* Norepinephrine controls astroglial responsiveness to local circuit activity. *Neuron* **82**, 1263–1270 (2014).
16. Srinivasan, R. *et al.* Ca^{2+} signaling in astrocytes from *Ip3r2^{-/-}* mice in brain slices and during startle responses *in vivo*. *Nat. Neurosci.* **18**, 708–717 (2015).
17. Zhou, Y., Cameron, S., Chang, W.-T. & Rao, Y. Control of directional change after mechanical stimulation in *Drosophila*. *Mol. Brain* **5**, 39–52 (2012).
18. Shigetomi, E., Jackson-Weaver, O., Huckstepp, R. T., O'Dell, T. J. & Khakh, B. S. TRPA1 channels are regulators of astrocyte basal calcium levels and long-term potentiation via constitutive D-serine release. *J. Neurosci.* **33**, 10143–10153 (2013).
19. Duffy, S. & MacVicar, B. A. Adrenergic calcium signaling in astrocyte networks within the hippocampal slice. *J. Neurosci.* **15**, 5535–5550 (1995).
20. Salm, A. K. & McCarthy, K. D. Norepinephrine-evoked calcium transients in cultured cerebral type 1 astroglia. *Glia* **3**, 529–538 (1990).
21. Vömel, M. & Wegener, C. Neuroarchitecture of aminergic systems in the larval ventral ganglion of *Drosophila melanogaster*. *PLoS One* **3**, e1848 (2008).
22. Nitabach, M. N., Sheeba, V., Vera, D. A., Blau, J. & Holmes, T. C. Membrane electrical excitability is necessary for the free-running larval *Drosophila* circadian clock. *J. Neurobiol.* **62**, 1–13 (2005).
23. Saudou, F., Amlaiky, N., Plassat, J. L., Borrelli, E. & Hen, R. Cloning and characterization of a *Drosophila* tyramine receptor. *EMBO J.* **9**, 3611–3617 (1990).
24. Robb, S. *et al.* Agonist-specific coupling of a cloned *Drosophila* octopamine/tyramine receptor to multiple second messenger systems. *EMBO J.* **13**, 1325–1330 (1994).
25. Kutsukake, M., Komatsu, A., Yamamoto, D. & Ishiwa-Chigusa, S. A tyramine receptor gene mutation causes a defective olfactory behavior in *Drosophila melanogaster*. *Gene* **245**, 31–42 (2000).
26. Burke, C. J. *et al.* Layered reward signalling through octopamine and dopamine in *Drosophila*. *Nature* **492**, 433–437 (2012).
27. Newman, E. A. Glial cell inhibition of neurons by release of ATP. *J. Neurosci.* **23**, 1659–1666 (2003).
28. Lalo, U., Rasooli-Nejad, S. & Pankratov, Y. Exocytosis of gliotransmitters from cortical astrocytes: implications for synaptic plasticity and aging. *Biochem. Soc. Trans.* **42**, 1275–1281 (2014).
29. Brody, T. & Cravchik, A. *Drosophila melanogaster* G protein-coupled receptors. *J. Cell Biol.* **150**, F83–F88 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank our colleagues, the Vienna *Drosophila* RNAi Center and the Bloomington Stock Center for providing fly stocks, members of the Freeman laboratory for comments on the manuscript and A. Sheehan for generating the *wtrw::gfp* construct. This work was supported by NINDS grant R01 NS053538 (to M.R.F.). During the period of this study M.R.F. was an Investigator with the Howard Hughes Medical Institute.

Author Contributions Z.M. and M.R.F. designed experiments. Z.M. performed all experiments. T.S. provided *alrm-LexA::GAD* transgenic flies. D.E.B. provided unpublished data on norepinephrine-mediated activation of mammalian astrocytes and input that helped guide the course of the study. Z.M. and M.R.F. wrote the manuscript with editing by T.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.F. (freemmar@ohsu.edu).

Reviewer Information Nature thanks L. Luo, B. MacVicar, L. Voss hall and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Fly stocks and husbandry. Flies were cultured in standard cornmeal food at 25°C in 12 h–12 h light–dark cycles. Fly stocks include: *w¹¹¹⁸*, *alrm-Gal4*, *alrm-LexA::GAD*, *UAS-wtrw-gfp*, *LexAop-R-GECO1* (Bloomington, 52224), *tdc2-Gal4* (Bloomington, 9313), *20XUAS-IVS-GCaMP6s* (Bloomington, 42746), *13XLexAop2-IVS-GCaMP6s-p10* (Bloomington, 44274), *UAS-Kir2.1-EGFP* (Bloomington, 6595), *UAS-Oct-TyrR^{RNAi}* (Bloomington, 28332), *th-Gal4* (Bloomington, 8848), *UAS-wtrw^{VDRC}* (VDRC, 107423), *Oct-TyrR^{homo}* (Kyoto, 109038), *UAS-shi^{DN}* (Bloomington, 5822), *UAS-TrpA1* (Bloomington, 26263), *UAS-NpHR* (Kyoto, 117008), *UAS-CsChrimson* (Bloomington, 55135), *AdoR^{KG03964ex}* (Bloomington, 30868), *tdc2-LexA::p65* (ref. 26), *UAS-wtrw^{2.23}* (ref. 13), *wtrw^{ex}* (ref. 30), *tβh^{nM18}* (ref. 31), *tdc2^{RO54}* (ref. 32), *th-LexA::p65* (ref. 33), *tsh-Gal80* (ref. 34).

Chemotaxis and light avoidance. These two assays were performed using protocols described previously with minor modifications³⁵. Briefly, pools of ~100 3rd instar larvae (108–120 h after egg lay) were allowed to move freely for 5 min on Petri dishes with according settings for chemotaxis assay or phototaxis assay, then the numbers of larvae in odorant versus vehicle circles or in light versus dark quadrants were scored, respectively. Response indices were calculated as: $RI_{chemotaxis} = (N_{odor} - N_{vehicle}) / (N_{odor} + N_{vehicle})$; $RI_{light\ avoidance} = (N_{dark} - N_{light}) / (N_{dark} + N_{light})$. IAA was diluted to 1:5,000 in mineral oil, and pure mineral oil was used as the control. Light intensity in phototaxis assays was 2,400 lx white light (5,400 K). Thermogenetic manipulation of dopaminergic neurons was accomplished by transferring 3rd instar larvae from 18°C to 25°C to allow activation of TrpA1 for 24 h before testing. For chemotaxis assays related to D1-like or D2-like dopamine receptor inhibitors, 3rd instar larvae were fed on cornmeal food mixed with inhibitors for 24 h before testing.

Gentle touch assay. We struck early 3rd instar larvae (84–96 h after egg lay) in the thoracic segments with a hair while they were moving straight. No response, a stop, head retraction and turn were grouped into type I responses, and initiation of at least one single full body retraction or multiple full body retractions were categorized as type II reversal responses.

Larval motility. Third instar larvae (108–120 h after egg lay) were videotaped and the maximal velocity was calculated by dividing the travel distance measured when larvae moved straight without pause by the time taken.

Molecular biology and transgenic flies. *pUAST-wtrw::gfp* was made by cloning full length *wtrw* tagged with *gfp* at its C-terminal into *pUAST*. Transgenic flies were obtained by standard germline injection (BestGene Inc.).

Chemicals, buffers and antibodies. For live preparation studies, all chemicals were diluted in live imaging buffer (110 mM NaCl, 5.4 mM KCl, 1.2 mM CaCl₂, 0.8 mM MgCl₂, 10 mM D-glucose, 10 mM HEPES, pH adjusted to 7.2), including 1 μM tetrodotoxin (TTX), 0.1 mM lanthanum chloride (LaCl₃), 2.5 mM Oct, 2.5 mM glutamate, 2.5 mM acetylcholine, 2.5 mM GABA, 150 μM AITC, 10 μM SCH-442416, 1 mM terazosin and 0.5 mM Tyr unless otherwise stated. The concentration of SCH-23390 (dopamine D1-like receptor antagonist) was 0.2 mg/ml; eticlopride (dopamine D2-like receptor antagonist) was 0.2 mg/ml; all-*trans* retinal was 10 mM. Rabbit anti-GAT antibody was used at 1:5,000 and mouse anti-nc82 antibody was used at 1:50 (DSHB).

Live imaging and data analysis. First instar larvae were immersed in halocarbon oil 27 and immobilized between coverslips and slides to image the activity of neurons and astrocytes through the cuticle in intact larvae. Third instar wandering stage larvae were used to dissect the CNS. Briefly, intact CNSs were transferred to a silicone filled dish after dissection and incubated with 100 μl live imaging buffer. Samples were allowed to sit at room temperature for 15 min before imaging. To examine the effects of LaCl₃ or tetrodotoxin on calcium transients, samples were

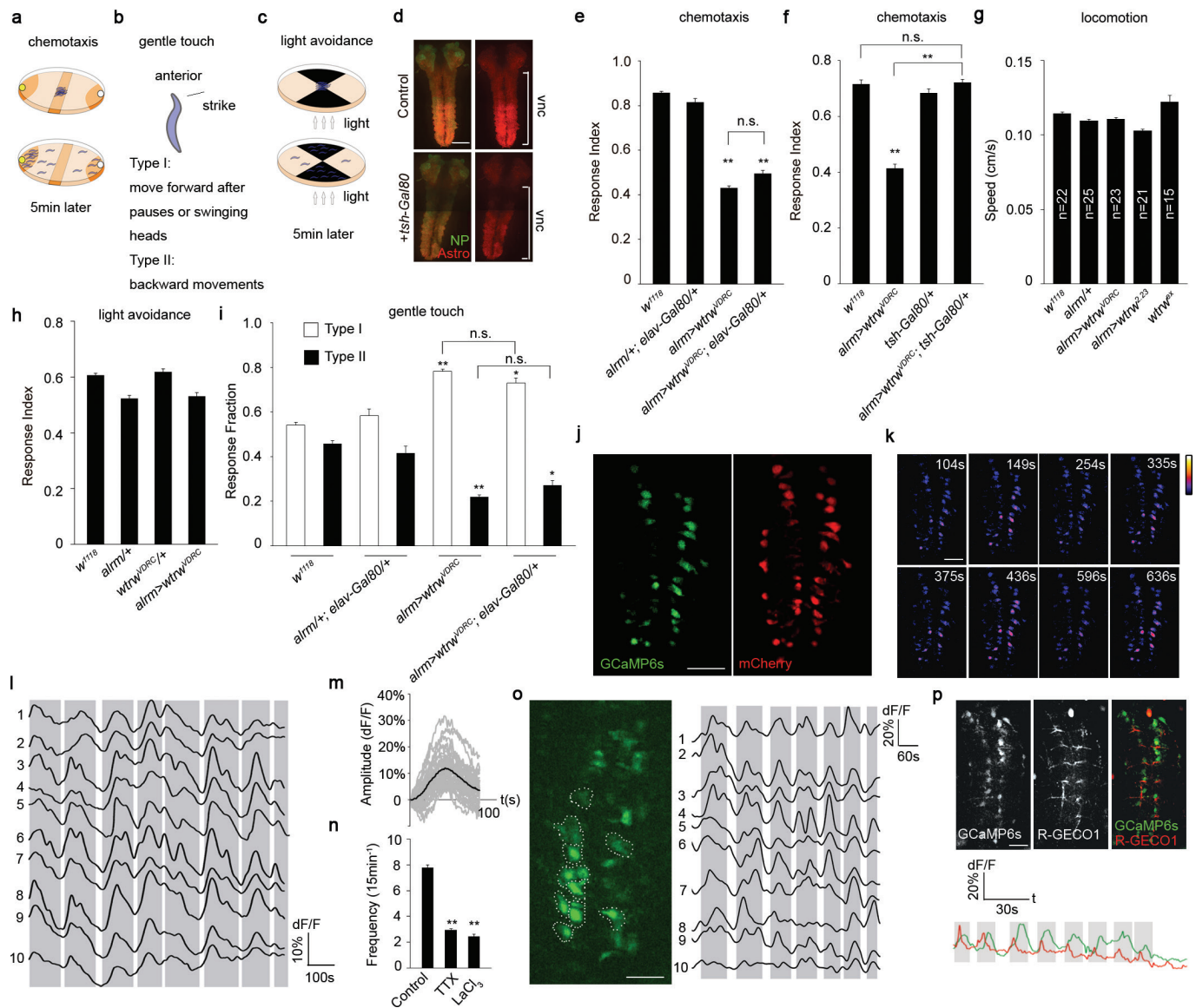
pre-incubated in buffer containing chemicals for 15 min before imaging. Movies were acquired by using Volocity software on a spinning disc microscope equipped with a 20× (dissected CNS) or 40× (intact larvae) water immersion objective at 2 frames per s. Changes in GCaMP intensity of ROIs were calculated and expressed as $(F_{t1...n} - F)/F$ (first 10 frames were averaged as *F*), which then were plotted over time. Plotted traces were semi-automatically analysed by the multiple peak fitting function of Igor Pro (WaveMetrics Inc.) to obtain the numbers and amplitudes of peaks.

Bath application of drugs. Samples were prepared as described above and incubated with imaging buffer (with/without 1 μM TTX). At ~240 s (astrocytes) or ~300 s (neurons) after imaging, 100 μl live imaging buffer containing 2× final concentrations of chemicals (with/without 1 μM TTX) were directly added to the medium, and imaging continued. The GCaMP6s intensity of ROIs was plotted and analysed as mentioned above. For terazosin experiments we first examined preparations for 6 min (control, pre-exposure window). Then terazosin was perfused for 3 min, and we then imaged for an additional 6 min (exposed window). Data points of 6–9 min that represented the perfusion window were omitted from the graph.

Optogenetic activation and inactivation of neurons. For CsChrimson activation of olfactory neurons, larvae were grown on cornmeal food daily topped with 100 μl all-*trans* retinal for 3 days. Early 3rd instar larvae were immobilized between coverslips and slides in halocarbon oil 27 and the activity of Tdc2⁺ neurons reflected by GCaMP6s was continuously monitored except when red light (640 nm, 3.41 mW/mm²) was delivered through an epifluorescence light source. Halorhodopsin-mediated inhibition of Tdc2⁺ neurons was accomplished by growing larvae on cornmeal food topped daily with 100 μl all-*trans* retinal for 4–5 days. Larval brains were dissected from wandering 3rd instar larvae and prepared as described above for imaging somatic calcium transients in astrocytes. GCaMP6s signals were recorded for 6 min, then halorhodopsin was activated (within ~30 s) by alternating delivery of 561 nm (0.26 mW/mm², 3 s, for photoinhibition) and blue light 488 nm (500 ms, for imaging GCaMP6s-labelled astrocytes) for another 6 min. Data points from two consecutive movies were then assembled together to show changes in somatic calcium transients.

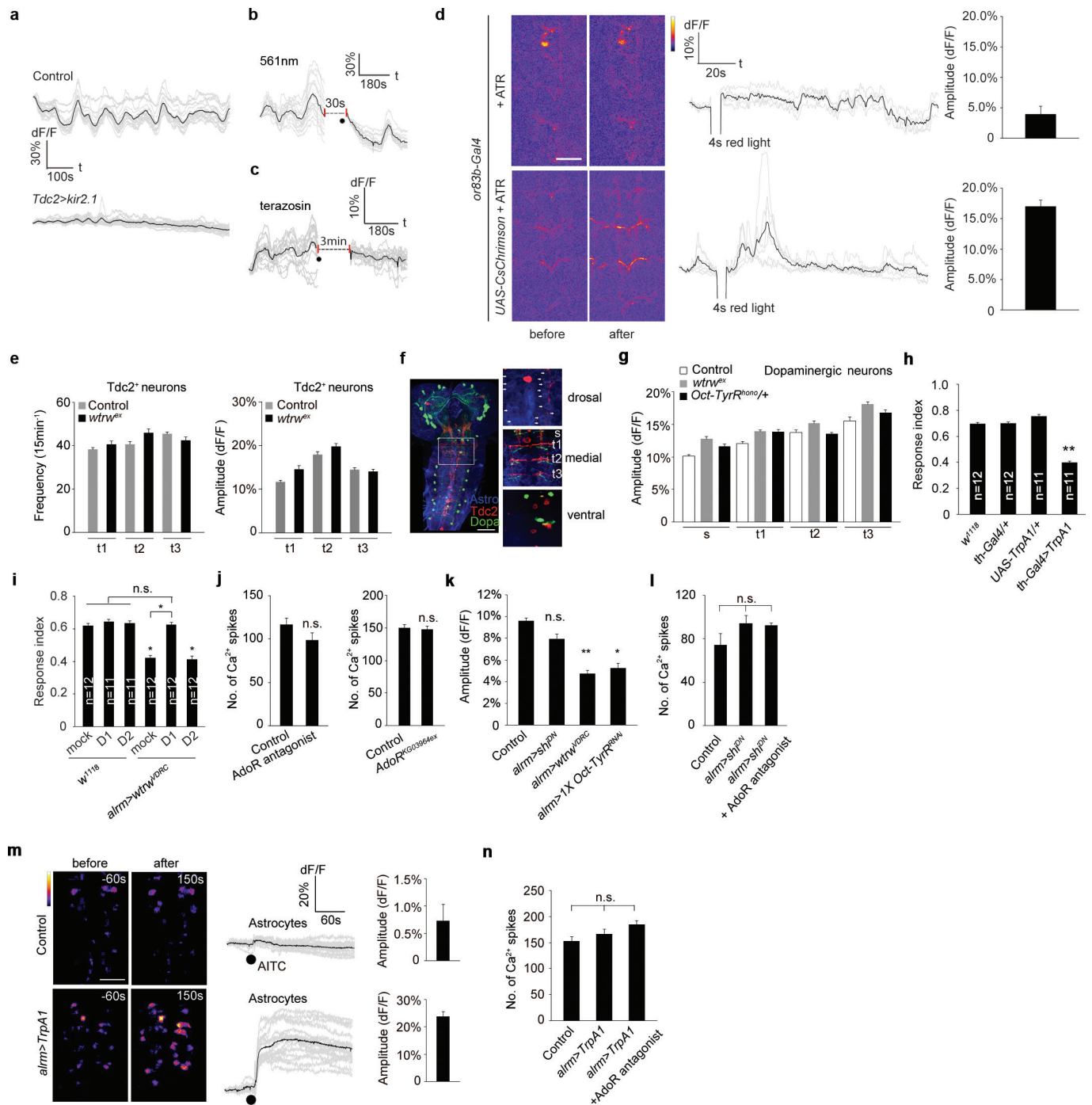
Statistical analysis. No statistical methods were used to predetermine sample size. All *n* numbers represent biological replicates. Data were pooled from 2–3 independent experiments. The experiments were not randomized or blinded. The Shapiro–Wilk test (normal distribution if *P* > 0.05) was used to determine the normality of data. Statistical comparisons were performed using one-way ANOVA followed by Tukey's *post hoc* test or paired Student's *t*-test (two-tailed). Non-normally distributed data were compared using Wilcoxon and Mann–Whitney tests followed by Bonferroni–Holm *post hoc* test. *P* < 0.05 was considered significant. All the data in bar graphs are expressed as mean ± s.e.m.

- Kim, S. H. *et al.* *Drosophila* TRPA1 channel mediates chemical avoidance in gustatory receptor neurons. *Proc. Natl Acad. Sci. USA* **107**, 8440–8445 (2010).
- Monastirioti, M., Linn, C. E., Jr & White, K. Characterization of *Drosophila* tyramine β-hydroxylase gene and isolation of mutant flies lacking octopamine. *J. Neurosci.* **16**, 3900–3911 (1996).
- Cole, S. H. *et al.* Two functional but noncomplementing *Drosophila* tyrosine decarboxylase genes: distinct roles for neural tyramine and octopamine in female fertility. *J. Biol. Chem.* **280**, 14948–14955 (2005).
- Galili, D. S. *et al.* Converging circuits mediate temperature and shock aversive olfactory conditioning in *Drosophila*. *Curr. Biol.* **24**, 1712–1722 (2014).
- Clyne, J. D. & Miesenböck, G. Sex-specific control and tuning of the pattern generator for courtship song in *Drosophila*. *Cell* **133**, 354–363 (2008).
- Lilly, M. & Carlson, J. smellblind: a gene required for *Drosophila* olfaction. *Genetics* **124**, 293–302 (1990).



Extended Data Figure 1 | Synchronous somatic Ca^{2+} transients in *Drosophila* astrocytes. **a–c**, Brief diagrams of behavioural tests. **d**, *tsh-Gal80* suppression of *alm-Gal4* activity. Nc82, neuropil (NP). Astrocytes (Astro), *alm > myr::tdTomato*. Vnc, ventral nerve cord. Scale bar, 50 μm . **e, f**, Chemotaxis assay ($n = 12$). **g**, Locomotion assay (n listed). **h**, Light avoidance assay ($n = 12$). **i**, Gentle touch assay ($n = 24$). **j**, GCaMP6s and mCherry expression in astrocytes. Scale bar, 50 μm . **k**, Representative pseudocoloured images of four continuous Ca^{2+} transients. Scale bar, 50 μm . **l**, Traces of normalized GCaMP6s intensity over mCherry of 10 individual astrocytes in 15-min live imaging windows. **m**, Averaged

traces of individual somatic Ca^{2+} transients. **n**, Somatic Ca^{2+} transients in astrocytes treated with TTX and LaCl_3 ($n = 10$, 160 cells). **o**, GCaMP6s expression in astrocytes and traces of 10 individual astrocytes from an intact larva. Scale bar, 20 μm . Grey bars (**l, o**) represent population rise or fall in GCaMP6s signals. **p**, GCaMP6s-labelled astrocytes and R-GECO1-labelled Tdc2⁺ neurites and their averaged traces in an intact larva. Scale bar, 20 μm . * $P < 0.05$, ** $P < 0.01$; NS, not significant; error bars, s.e.m. Wilcoxon and Mann–Whitney tests followed by Bonferroni–Holm *post hoc* test (**e, n**), one-way ANOVA followed by Tukey’s *post hoc* test (**f–i**).



Extended Data Figure 2 | Somatic Ca^{2+} transients in astrocytes inhibit the activity of dopaminergic neurons. a–c, Representative traces of astrocyte Ca^{2+} transients with blockade of Tyr and Oct signalling.

d, Stimulation of olfactory neurons activates *Tdc2*⁺ neurons ($n=3-4$). Scale bar, 25 μ m. e, Activity of *Tdc2*⁺ neurons is not altered in *wtrw* mutants ($n=8$, 48 neurites). f, Astrocytes, *Tdc2*⁺ neurons and dopaminergic neurons in larval CNS. Dorsal (arrows point to astrocyte somas labelled with anti-GAT antibody), medial (neurites intermingled with ramified processes of astrocytes for monitoring activity are labelled) and ventral (cell bodies of *tdc2>myr::tdTomato*, *Gal4/UAS* and *th>GCaMP6s* LexA/LexAop dopaminergic neurons) images from

the boxed region are shown (right). s, subesophageal. t, thoracic. Scale bar, 50 μ m. g, Amplitude of Ca^{2+} spikes in dopaminergic neurons ($n=10$, 80 neurites). h, i, Chemotaxis assay (n listed). j, Number of Ca^{2+} spikes of dopaminergic neurons ($n=6$, 48 neurites). k, Responses of astrocytes to Tyr (0.5 mM) in the presence of TTX ($n=6$, 96 cells total). l, Number of Ca^{2+} spikes in dopaminergic neurons ($n=6$, 48 neurites). m, AITC induces Ca^{2+} influx into astrocytes expressing *TrpA1* ($n=5$, 80 cells). Scale bar, 50 μ m. n, Number of Ca^{2+} spikes in dopaminergic neurons ($n=6$, 48 neurites). * $P < 0.05$, ** $P < 0.01$; NS, not significant; error bars, s.e.m. One-way ANOVA followed by Tukey's *post hoc* test.

Transcription of the non-coding RNA upperhand controls *Hand2* expression and heart development

Kelly M. Anderson^{1,2}, Douglas M. Anderson^{1,2}, John R. McAnally^{1,2}, John M. Shelton³, Rhonda Bassel-Duby^{1,2} & Eric N. Olson^{1,2}

HAND2 is an ancestral regulator of heart development and one of four transcription factors that control the reprogramming of fibroblasts into cardiomyocytes^{1–4}. Deletion of *Hand2* in mice results in right ventricle hypoplasia and embryonic lethality^{1,5}. *Hand2* expression is tightly regulated by upstream enhancers^{6,7} that reside within a super-enhancer delineated by histone H3 acetyl Lys27 (H3K27ac) modifications⁸. Here we show that transcription of a *Hand2*-associated long non-coding RNA, which we named upperhand (*Uph*), is required to maintain the super-enhancer signature and elongation of RNA polymerase II through the *Hand2* enhancer locus. Blockade of *Uph* transcription, but not knockdown of the mature transcript, abolished *Hand2* expression, causing right ventricular hypoplasia and embryonic lethality in mice. Given the substantial number of uncharacterized promoter-associated long non-coding RNAs encoded by the mammalian genome⁹, the *Uph*–*Hand2* regulatory partnership offers a mechanism by which divergent non-coding transcription can establish a permissive chromatin environment.

In all well-annotated mammalian species, long non-coding RNA (lncRNA) transcripts with a similar intron–exon organization are transcribed upstream of the *Hand2* locus (Extended Data Fig. 1a, b). We refer to these transcripts as ‘upperhand’ (*Uph*), although the human transcripts have been referred to as *DEIN* (differentially expressed in neuroblastoma, also known as *HAND2-AS1*)¹⁰. Rapid amplification of cDNA ends (RACE) analysis of *Uph* revealed several alternatively spliced transcripts in the mouse heart (Extended Data Fig. 1c). The major mouse *Uph* transcript is 770 nucleotides long, contains six exons, a polyadenylation sequence and shares a bidirectional promoter with *Hand2* (ref. 10; Fig. 1a). The mouse *Uph* locus encompasses ~16.5 kilobases and contains two enhancers that direct *Hand2* expression in the heart and branchial arches^{6,7}, as well as histone acetylation marks that delineate a cardiac super-enhancer in this locus⁸ (Fig. 1a).

Similar to most characterized lncRNAs¹¹, the *Uph* nucleotide sequence is not well-conserved, with only ~56% homology between mouse and human, and little conservation across other species (Extended Data Fig. 1a). However, *Uph* orthologues share a promoter and contain the conserved *Hand2*-associated cardiac and branchial arch enhancers within their second introns (Extended Data Fig. 1b).

Promoter-associated lncRNAs are often positively correlated with transcription of their protein-coding neighbour^{12,13}. Indeed, whole-mount *in situ* hybridization for *Uph* in mouse embryos revealed expression in the heart, distal branchial arches and the developing limb bud at embryonic day (E)10.5, overlapping with *Hand2* expression (Extended Data Fig. 1d). *Uph* expression was strong in the heart and continued to mirror *Hand2* at later fetal stages (Fig. 1b). Northern blot analysis across several adult tissues revealed that *Uph* transcript levels were highest in the heart (Extended Data Fig. 1e).

Cell fractionation of mouse neonatal cardiomyocytes showed *Uph* transcripts in all fractions but enrichment in the cytoplasm, similar to 18S RNA (Extended Data Fig. 2a). The distribution of *Uph* differed

from nuclear lncRNAs such as *Malat1*, which associates with chromatin (Extended Data Fig. 2a). Previously, we reported that some annotated lncRNAs can encode micropeptides^{14,15}. However, *Uph* does not contain a conserved open reading frame and *in vitro* transcription and translation of *Uph* did not produce any detectable peptides (Extended Data Fig. 2b). Together, these data reveal that *Uph* is a bona fide lncRNA that is co-expressed in a temporal and tissue-specific pattern with the essential cardiac transcription factor *Hand2*.

To examine *Uph* function *in vivo*, we used transcription activator-like effector nucleases (TALENs) to insert a triple polyadenylation sequence into exon 2 of the *Uph* locus in C57BL/6 mouse zygotes (Extended Data Fig. 3a). To avoid potential regulatory elements required for *Hand2* expression, we inserted the triple polyadenylation cassette into a region with low sequence conservation. Furthermore, we generated a separate knock-in allele using a heterologous DNA sequence (tdTomato (tdTO) lacking a polyadenylation sequence) inserted into the same *Uph* locus (Extended Data Fig. 3b). Correct targeting of the triple polyadenylation and tdTO sequences was verified by Southern blot analysis (Extended Data Fig. 3c, d).

Mice homozygous for the *Uph* tdTO allele (*Uph*^{tdTO/tdTO}) were born at expected Mendelian ratios and showed no morphological defects or

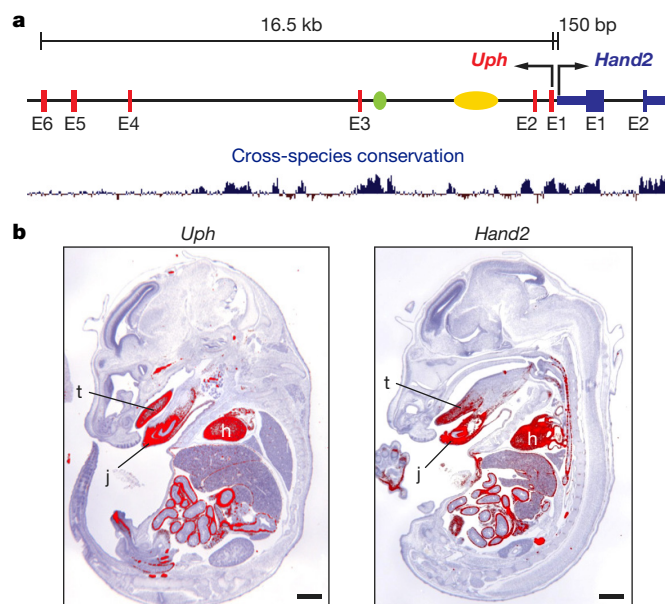


Figure 1 | Co-expression of upperhand (*Uph*) and *Hand2*. **a**, The mouse *Uph*–*Hand2* locus. *Uph* is transcribed 150 bp upstream of *Hand2* and contains two *Hand2* enhancers (branchial arch enhancer (green) and cardiac enhancer (yellow)) within one of its introns. Bottom, mammalian sequence conservation (blue). **b**, Section *in situ* hybridization of E15.5 mouse embryos for either *Uph* (left) or *Hand2* (right). Expression was detected in heart (h), tongue (t) and jaw (j). Scale bars, 1 mm.

¹Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ²Hamon Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ³Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

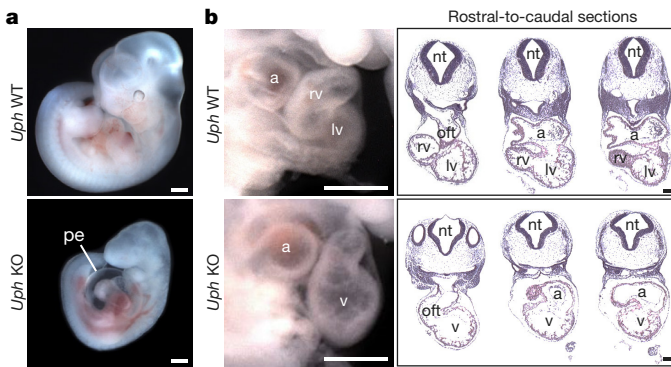


Figure 2 | Cardiac defects of *Uph* KO embryos. **a**, At E10.5, *Uph* KO embryos have pericardial effusion (pe) and appear smaller than wild-type (WT) littermate embryos. **b**, At E10.5, wild-type embryos (top) have distinct right and left ventricular (rv and lv, respectively) chambers, whereas *Uph* KO embryos (bottom) have a single ventricle (v). This abnormality is highlighted in rostral-to-caudal sections through the heart (left to right). a, atria; nt, neural tube; oft, outflow tract. Scale bars, 500 μ m (a) and 100 μ m (b).

notable changes in *Uph* or *Hand2* expression (Extended Data Fig. 3e, f). Thus, insertion of a heterologous DNA sequence into exon 2 of the *Uph* locus is not sufficient to disrupt normal *Hand2* expression. However, from 31 pups born of heterozygous *Uph* (*Uph*^{+/-}) intercrosses, no *Uph*-knockout (KO) offspring survived, suggesting that *Uph* transcription is required for embryonic survival.

To pinpoint the embryonic death of *Uph* KO embryos, we collected embryos from timed pregnancies of *Uph*^{+/-} matings. *Uph* KO embryos in a pure C57BL/6 background were present at approximate Mendelian ratios before E9.5 (20.6%; 13 out of 63 embryos collected), and showed no obvious size or morphological differences relative to wild-type littermates. However, by E10.5, the *Uph* KO embryos had a large pericardial effusion surrounding the heart and were growth restricted (Fig. 2a). Quantitative real-time PCR (qPCR) showed that *Uph* transcripts were reduced by 97% in *Uph* KO compared to wild-type hearts at E10.5, confirming successful termination of *Uph* transcription (Extended Data Fig. 3g). Histological analyses revealed that hearts from *Uph* KO embryos failed to develop a right ventricular chamber. Instead, the outflow tract and atrial chamber connected to a single left-sided ventricular chamber (Fig. 2b), a cardiac phenotype reminiscent of *Hand2* KO embryos.

Because *Uph* KO embryos displayed a similar cardiac phenotype to *Hand2* KO embryos, we analysed the expression of *Hand2* and other cardiac transcription factors in *Uph* KO embryos using

in situ hybridization. Notably, *Hand2* expression was absent in the atria, ventricle and outflow tract of *Uph* KO embryos, but remained detectable, albeit at reduced levels, in the branchial arches and limb buds (Fig. 3a, Extended Data Fig. 4a). Consistent with the persistent expression of *Hand2* in the branchial arches, *Uph* KO embryos lacked the vascular defects in the aortic arch arteries and the aortic sac typically seen in *Hand2* mutant embryos^{1,5} (Extended Data Fig. 4b). Genes encoding other cardiac transcription factors essential for cardiac morphogenesis, namely *Nkx2-5*, *Gata4* and the related *Hand1*, were not considerably changed in *Uph* KO embryos (Fig. 3b, Extended Data Fig. 4a), demonstrating the specificity of *Uph* transcription in regulating *Hand2* expression.

To determine whether *Uph* functions in *cis* or *trans* to regulate *Hand2*, we intercrossed mice carrying a *Hand2* KO allele with mice carrying the *Uph* KO allele, to generate compound heterozygous (*Uph*^{+/-} *Hand2*^{+/-}) embryos. Notably, *Uph*^{+/-} *Hand2*^{+/-} embryos recapitulated the *Uph* KO phenotype, displaying pericardial effusion and embryonic lethality by E10.5 (Fig. 4a, Extended Data Fig. 5a). While *Uph* expression was not reduced in compound heterozygotes, *Hand2* expression was decreased by greater than 90% (Extended Data Fig. 5b). These findings demonstrate that expression of *Hand2* in the heart is dependent on transcription of *Uph* in *cis*, which cannot be compensated for by *Uph* transcripts generated on the sister allele.

Several lncRNAs have been proposed to function by recruiting the trithorax chromatin-modifying complex to gene loci^{16,17}. To determine whether *Uph* interacts with the trithorax complex, we performed RNA immunoprecipitation using an antibody against the RNA-binding adaptor of the MLL-1 trithorax protein, WDR5. While the lncRNA HOTTIP formed a stable complex with WDR5 in the mouse HL-1 cardiomyocyte-like cell line¹⁸, we could not detect an interaction between MLL-1 and *Uph* (Extended Data Fig. 5c).

To determine whether the mature *Uph* RNA transcript is required for *Hand2* expression, we performed knockdown of *Uph* using GapmeR antisense oligonucleotides¹⁹ in HL-1 cells and in the neuroblastoma cell line Neuro2a, both of which express *Uph* and *Hand2* (Extended Data Fig. 6a). Knockdown of mature *Uph* transcripts by more than 90% using two different GapmeR oligonucleotides in both cell lines did not alter *Hand2* expression (Extended Data Fig. 6b, c). Because *Uph* is alternatively spliced, we performed qPCR across all *Uph* exon junctions (Extended Data Fig. 6d). We further verified that efficient knockdown was attained in the nuclear fraction of HL-1 cells (Extended Data Fig. 6e). Additionally, overexpression of the full-length *Uph* transcript in HL-1 cells using transient transfection had no effect on the levels of *Hand2* mRNA (Extended Data Fig. 6f). Together, these data suggest that, in contrast to premature transcriptional termination of *Uph* *in vivo*, the mature *Uph* transcript is not required for *Hand2* expression.

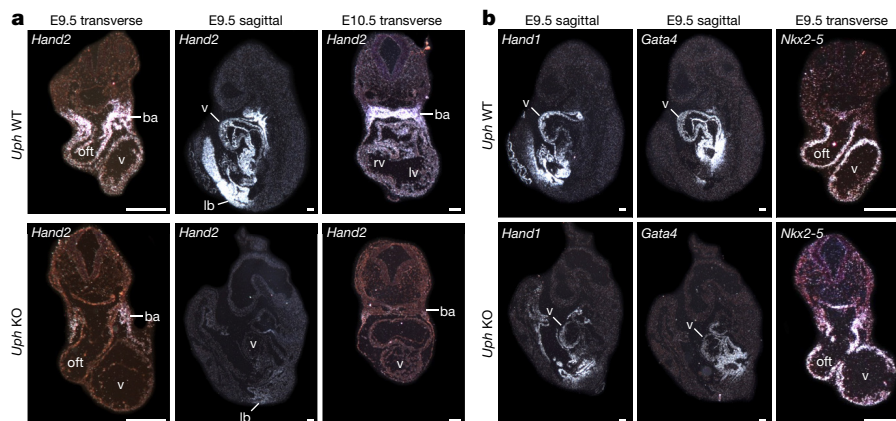
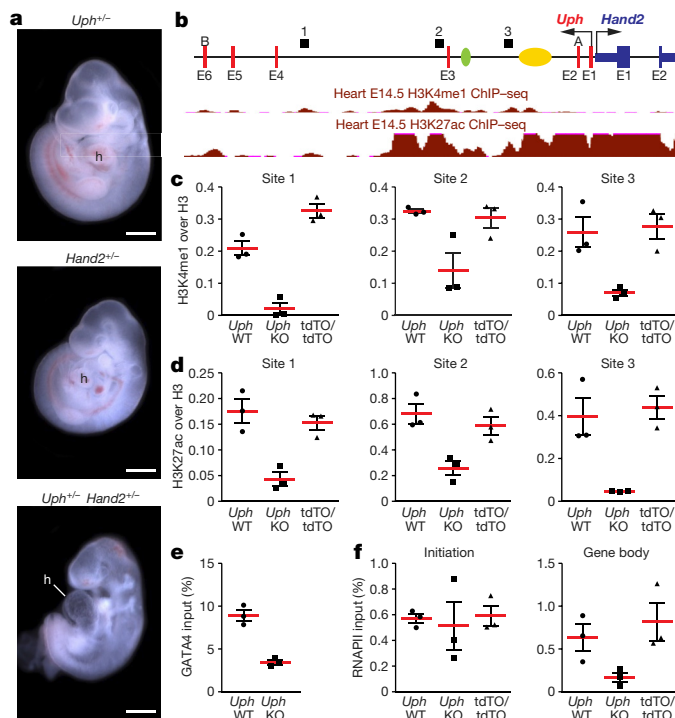


Figure 3 | Loss of cardiac *Hand2* expression in *Uph* KO embryos. **a**, *In situ* hybridizations of wild-type and *Uph* KO embryos at E9.5 (transverse and sagittal sections) and E10.5 (transverse) show loss of *Hand2* expression in the heart of *Uph* KO embryos, with detectable

expression in the branchial arches (ba) and limb bud (lb). **b**, *In situ* hybridizations of wild-type and *Uph* KO embryos at E9.5 (transverse and sagittal sections) using *Hand1*, *Gata4* or *Nkx2-5* probes. Scale bars, 100 μ m.



DNA loci with high regulatory capacity such as enhancers^{20,21}, polycomb response elements^{22,23} or locus control regions²⁴ are often associated with non-coding transcription and can function to maintain a permissive histone signature. In the developing heart, the super-enhancer at the *Hand2* locus is defined by H3K27ac (Fig. 4b). To test whether *Uph* transcription is required to maintain the enhancer signature in the *Uph*–*Hand2* locus, we performed chromatin immunoprecipitation (ChIP) on *Uph* KO and wild-type E10.0 hearts, using H3K4me1 and H3K27ac active enhancer marker antibodies. We found that the H3K4me1 and H3K27ac peaks normally established at the *Uph*–*Hand2* locus were reduced in *Uph* KO hearts (Fig. 4c, d).

The GATA4 transcription factor is required for activation of the HAND2 cardiac enhancer in the *Uph* locus, and participates in establishing active H3K27ac marks^{7,25}. We observed a reduction in GATA4 binding to the *Hand2* cardiac enhancer in *Uph* KO hearts, a finding consistent with the reduced H3K27ac enhancer marks in the *Uph* locus (Fig. 4e). We observed no differences in H3K4me1 or H3K27ac levels at an active enhancer in the promoter region of *Nkx2-5*, no differences in GATA4 binding to the *Nkx2-5* promoter, and no differences in

H3K27me3 repressive marks between wild-type and *Uph* KO embryos at the *Uph*–*Hand2* locus (Extended Data Fig. 7b–f).

Enhancer-associated histone acetylation is necessary for gene expression by promoting RNA polymerase II (RNAPII) elongation at gene promoters²⁶. To measure the levels of RNAPII at the *Hand2* locus, we performed ChIP on wild-type and *Uph* KO E10.0 hearts using the Ser2-phosphorylated RNAPII antibody, a marker of RNAPII elongation, at the *Hand2* locus. Although there was no difference in RNAPII recruitment to the *Hand2* transcriptional start site, RNAPII binding to the body of the *Hand2* locus was decreased in *Uph* KO hearts (Fig. 4f). We found no differences in Ser2-phosphorylated RNAPII binding to the *Nkx2-5* gene body between genotypes (Extended Data Fig. 7g). These findings suggest that the loss of GATA4 and histone acetylation marks in the *Uph*–*Hand2* locus of *Uph* KO hearts prevents RNAPII elongation within the *Hand2* locus.

Here we define the *in vivo* function of *Uph*, a cardiac-enriched lncRNA co-transcribed bi-directionally with the cardiac transcription factor *Hand2*. Termination of *Uph* transcription, but not of the mature transcript, resulted in the loss of *Hand2* expression in the heart and partially phenocopied the lethal cardiac defects observed in *Hand2* KO embryos. Compound *Uph*^{+/-} *Hand2*^{+/-} heterozygosity causes embryonic lethality, suggesting that transcription of *Uph* is required *in cis* to promote *Hand2* expression. Termination of *Uph* transcription *in vivo* reduced GATA4 binding and the *Uph*–*Hand2* super-enhancer signature, resulting in RNAPII pausing and loss of *Hand2* expression in all tissues for which the expression of *Hand2* depends on enhancers located within the *Uph* locus. Future studies are needed to determine the role of *Uph* in the regulation of *Hand2* in non-cardiac tissues, which are currently precluded owing to early embryonic lethality of the global *Uph* KO. Given the importance of *Hand2* in the development of the palate, tongue, limbs, and sympathetic nervous system^{27–30}, *Uph* provides a potentially important non-coding regulatory mechanism for the temporal and spatial control of *Hand2* expression.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 December 2015; accepted 29 September 2016.

Published online 26 October 2016.

1. Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat. Genet.* **16**, 154–160 (1997).
2. Han, Z., Yi, P., Li, X. & Olson, E. N. Hand, an evolutionarily conserved bHLH transcription factor required for *Drosophila* cardiogenesis and hematopoiesis. *Development* **133**, 1175–1182 (2006).
3. Song, K. *et al.* Heart repair by reprogramming non-myocytes with cardiac transcription factors. *Nature* **485**, 599–604 (2012).
4. Conway, S. J., Firulli, B. & Firulli, A. B. A bHLH code for cardiac morphogenesis. *Pediatr. Cardiol.* **31**, 318–324 (2010).
5. Yamagishi, H., Olson, E. N. & Srivastava, D. The basic helix-loop-helix transcription factor, dHAND, is required for vascular development. *J. Clin. Invest.* **105**, 261–270 (2000).
6. Charité, J. *et al.* Role of Dlx6 in regulation of an endothelin-1-dependent, dHAND branchial arch enhancer. *Genes Dev.* **15**, 3039–3049 (2001).
7. McFadden, D. G. *et al.* A GATA-dependent right ventricular enhancer controls dHAND transcription in the developing heart. *Development* **127**, 5331–5341 (2000).
8. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
9. Lepoivre, C. *et al.* Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics* **14**, 914 (2013).
10. Voth, H. *et al.* Co-regulated expression of HAND2 and DEIN by a bidirectional promoter with asymmetrical activity in neuroblastoma. *BMC Mol. Biol.* **10**, 28 (2009).
11. Chodroff, R. A. *et al.* Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* **11**, R72 (2010).
12. Kambara, H. *et al.* Regulation of interferon-stimulated gene BST2 by a lncRNA transcribed from a shared bidirectional promoter. *Front. Immunol.* **5**, 676 (2015).
13. Uesaka, M. *et al.* Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals. *BMC Genomics* **15**, 35 (2014).

14. Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606 (2015).
15. Nelson, B. R. *et al.* A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* **351**, 271–275 (2016).
16. Grote, P. *et al.* The tissue-specific lncRNA *Fendrr* is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* **24**, 206–214 (2013).
17. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
18. Yang, Y. W. *et al.* Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *eLife* **3**, e02046 (2014).
19. Stein, C. A. *et al.* Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents. *Nucleic Acids Res.* **38**, e3 (2010).
20. Kaikkonen, M. U. *et al.* Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* **51**, 310–325 (2013).
21. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
22. Schmitt, S., Prestel, M. & Paro, R. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev.* **19**, 697–708 (2005).
23. Herzog, V. A. *et al.* A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. *Nat. Genet.* **46**, 973–981 (2014).
24. Ho, Y., Elefant, F., Liebhaber, S. A. & Cooke, N. E. Locus control region transcription plays an active role in long-range gene activation. *Mol. Cell* **23**, 365–375 (2006).
25. He, A. *et al.* Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat. Commun.* **5**, 4907 (2014).
26. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
27. Yanagisawa, H., Clouthier, D. E., Richardson, J. A., Charité, J. & Olson, E. N. Targeted deletion of a branchial arch-specific enhancer reveals a role of dHAND in craniofacial development. *Development* **130**, 1069–1078 (2003).
28. Barron, F. *et al.* Downregulation of *Dlx5* and *Dlx6* expression by *Hand2* is essential for initiation of tongue morphogenesis. *Development* **138**, 2249–2259 (2011).
29. Charité, J., McFadden, D. G. & Olson, E. N. The bHLH transcription factor dHAND controls Sonic hedgehog expression and establishment of the zone of polarizing activity during limb development. *Development* **127**, 2461–2470 (2000).
30. Lucas, M. E., Müller, F., Rüdiger, R., Henion, P. D. & Rohrer, H. The bHLH transcription factor *hand2* is essential for noradrenergic differentiation of sympathetic neurons. *Development* **133**, 4015–4024 (2006).

Acknowledgements We thank D. Tension for technical assistance and J. Cabrera for graphics. We thank the Histopathology Core Facility at University of Texas Southwestern for histology. This work was supported by grants from the National Institutes of Health (HL-077439, AR-067294, HL-130253, DK-099653 and U01-HL-100401), Fondation Leducq Networks of Excellence, Cancer Prevention and Research Institute of Texas and the Robert A. Welch Foundation (grant 1-0025 to E.N.O.), a pre-doctoral fellowship from the American Heart Association (14PRE19830031) to K.M.A. and a Muscular Dystrophy Association Development Grant (MDA377340) to D.M.A.

Author Contributions K.M.A., D.M.A. and E.N.O. designed experiments and analysed data. J.R.M. generated mutant mice from constructs designed and created by K.M.A. and D.M.A. K.M.A. performed qPCR, phenotypic analysis, dissections, ChIP, RNA immunoprecipitation, cardiomyocyte fractionation, Southern blot, and India ink injections. D.M.A. performed northern blot, RACE, *in vitro* transcription/translation, and GapmeR transfections. J.M.S. and K.M.A. performed *in situ* hybridization and generated images. K.M.A., D.M.A., E.N.O. and R.B.D. wrote and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.N.O. (eric.olson@utsouthwestern.edu).

Reviewer Information *Nature* thanks H. Chang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Animal models. All experimental procedures involving animals in this study were reviewed and approved by the University of Texas Southwestern Medical Center's Institutional Animal Care and Use Committee. Animals/embryos were allocated to experimental groups based on genotype and we did not use exclusion, randomization or blinding approaches. The sex of embryos used in these studies was not determined. In general, sample size was chosen to use the least number of animals/embryos to achieve statistical significance and no statistical methods were used to predetermine sample size.

In situ hybridization. Whole mount *in situ* hybridization was performed using digoxigenin-labelled antisense RNA probes specific to *Uph* and *Hand2*, using methods previously described³¹. Radioisotopic *in situ* hybridization studies on sections were performed as previously described³². Antisense RNA probe templates for *Uph* or *Hand2* were generated using mouse heart cDNA and subcloned into pCRII TOPO (Life Technologies). Primer sequences are listed in Extended Data Table 1.

Subcellular fractionation of mouse neonatal cardiomyocytes. Neonatal cardiomyocytes were collected from ~50 post-natal day 1 (P1) C57BL/6 mice using the Neomyt kit (Cellutron, nc-6031) and fractionated as previously described³³.

TALEN-mediated homologous recombination in mice. A TALEN pair specific for the *Uph* locus was designed using the ZiFiT Targeter Program (<http://zifit.partners.org/ZiFiT/Introduction.aspx>). Individual TALEN modules and donor vectors were constructed and assembled as previously described¹⁴. TALEN mRNAs and circular DNA donor plasmids were diluted to 25 ng μl^{-1} and 3 ng μl^{-1} , respectively, and co-injected into the nucleus and cytoplasm of one-cell stage C57BL/6 zygotes and transferred into pseudopregnant mice. Cloning and genotyping primer sequences are listed in Extended Data Table 1.

Histology. Embryos were fixed in 4% formaldehyde in PBS and processed for paraffin histology, sectioned and stained with haematoxylin and eosin using routine procedures. At least 5 embryos of each genotype were sectioned for all experiments and data reported are representative of all samples collected.

ChIP. ChIP was performed using the SimpleChIP Kit (Cell Signaling, 9004) with the following adjustments to the standard protocol. In brief, embryonic hearts were dissected from E10.0 embryos and flash frozen in liquid nitrogen. On average, 5 hearts from identical genotypes were pooled in a 1.5 ml tube before formaldehyde cross-linking. Sonication was performed using a Bioruptor (Diagenode). The following antibodies were used: 10 μl H3K4me1 (Abcam, ab8895), 10 μl histone H3 (Cell Signaling), 10 μl H3K27ac (Active Motif, 39133), 10 μl H3K27me3 (Active Motif, 39155), 5 μl RNA polymerase II CTD repeat YSPTSPS (phospho S2) (Abcam, ab5095), 5 μl of GATA4 (sc-1237X). qPCR primer sequences are listed in Extended Data Table 1.

GapmeR antisense oligonucleotide knockdown of *Uph* in HL-1 and Neuro2a cells. The mouse neuroblastoma cell line Neuro2a was purchased from ATCC

and authenticated by morphology. The cardiomyocyte-like cell line HL-1 was donated by the laboratory of W. Claycomb. Contamination from mycoplasma was not tested for in these cell lines. LNA longRNA GapmeR oligonucleotides specific for *Uph* were designed and purchased from Exiqon. HL-1 and Neuro2a cells were co-transfected with a plasmid encoding eGFP (CS2GFP, 0.3 ng μl^{-1} final) and 50 nM GapmeR antisense oligonucleotides using Lipofectamine2000 (Life Technologies), according to the manufacturer's recommended protocol. Two days after transfection, GFP-positive cells were FACS sorted and processed using TRIzol (Life Technologies) for downstream gene expression analyses. GapmeR antisense oligonucleotide sequences are listed in Extended Data Table 1.

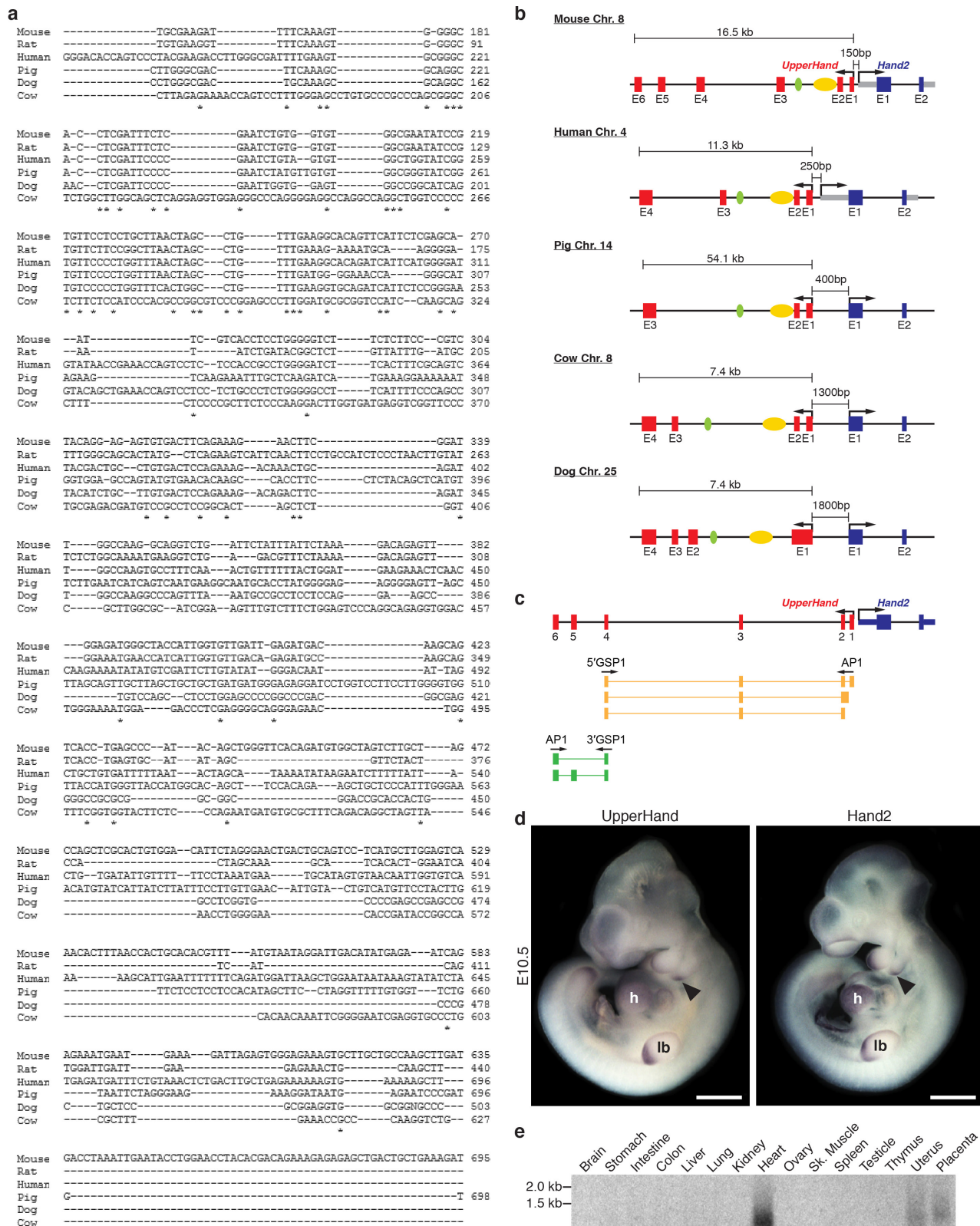
RACE PCR. The sequences of *Uph* RNA transcripts expressed in the mouse heart were determined using 5' and 3' RACE PCR using commercially available Marathon-Ready cDNA (Clontech). RACE PCR was performed according to manufacturer's recommended protocol, using primers specific to exon 4 of *Uph* and the marathon adaptor primer 1 (AP1). 5' and 3' *Uph* primer sequences are listed in Extended Data Table 1.

RNA immunoprecipitation. RNA-protein complexes from formaldehyde-cross-linked HL-1 cell lysates were immunoprecipitated using 2 μl IgG (Cell Signaling) or 2 μl WDR5 (Cell Signaling) antibodies. RNA purification and expression analyses were performed as previously described³⁴.

Statistics. Unless otherwise indicated in the figure legends, error bars represent s.e.m. of at least 3 biological replicates of each genotype or sample.

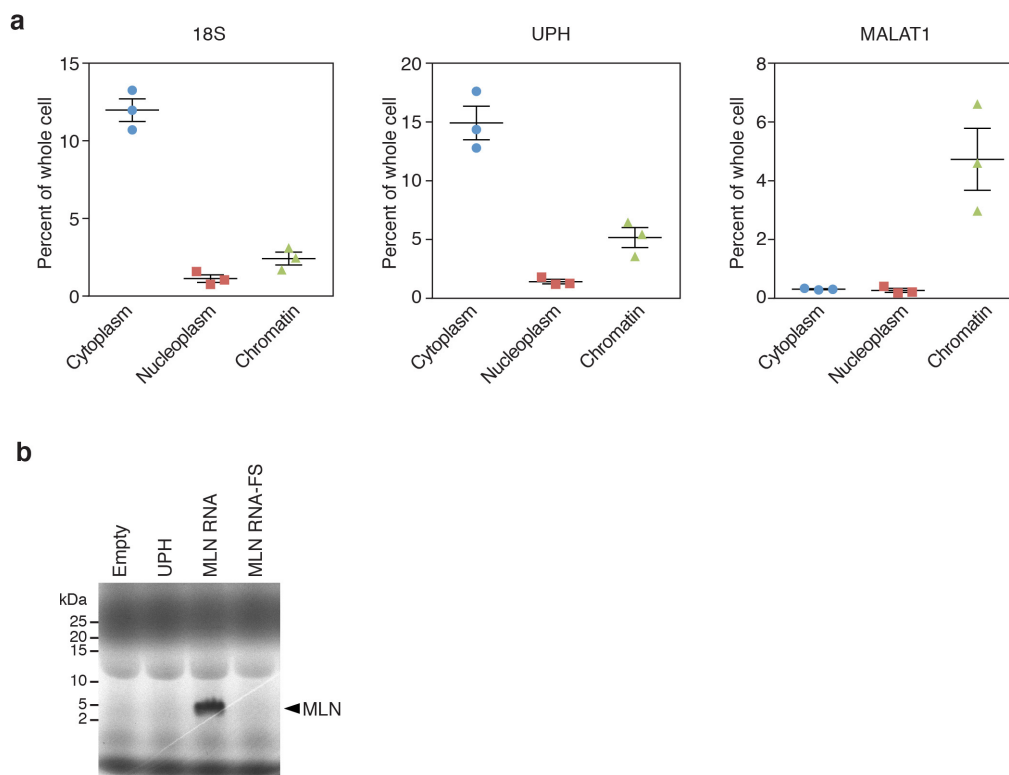
Data availability. Expressed sequence tag (EST) sequences for alignment of *Uph* transcripts were downloaded from GenBank: rat (BF567084), pig (DN117952), dog (DN434464) and cow (DN282558), or from RefSeq: human (NR_003679.2). Previously published ChIP-seq data used in Fig. 4b and Extended Data Fig. 7a, b are available under accession code GSE31039 (ref. 35). Source data for Extended Data Figs 1e, 2b, 3c, d are available in Supplementary Fig. 1, and all other data that support the findings of this study are available from the corresponding author on request.

- Anderson, D. M. *et al.* Mohawk is a novel homeobox gene expressed in the developing mouse embryo. *Dev. Dyn.* **235**, 792–801 (2006).
- Shelton, J. M., Lee, M. H., Richardson, J. A. & Patel, S. B. Microsomal triglyceride transfer protein expression during mouse development. *J. Lipid Res.* **41**, 532–537 (2000).
- Gagnon, K. T., Li, L., Janowski, B. A. & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protocols* **9**, 2045–2060 (2014).
- Conrad, N. K. Chapter 15. Co-immunoprecipitation techniques for assessing RNA-protein interactions in vivo. *Methods Enzymol.* **449**, 317–342 (2008).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).



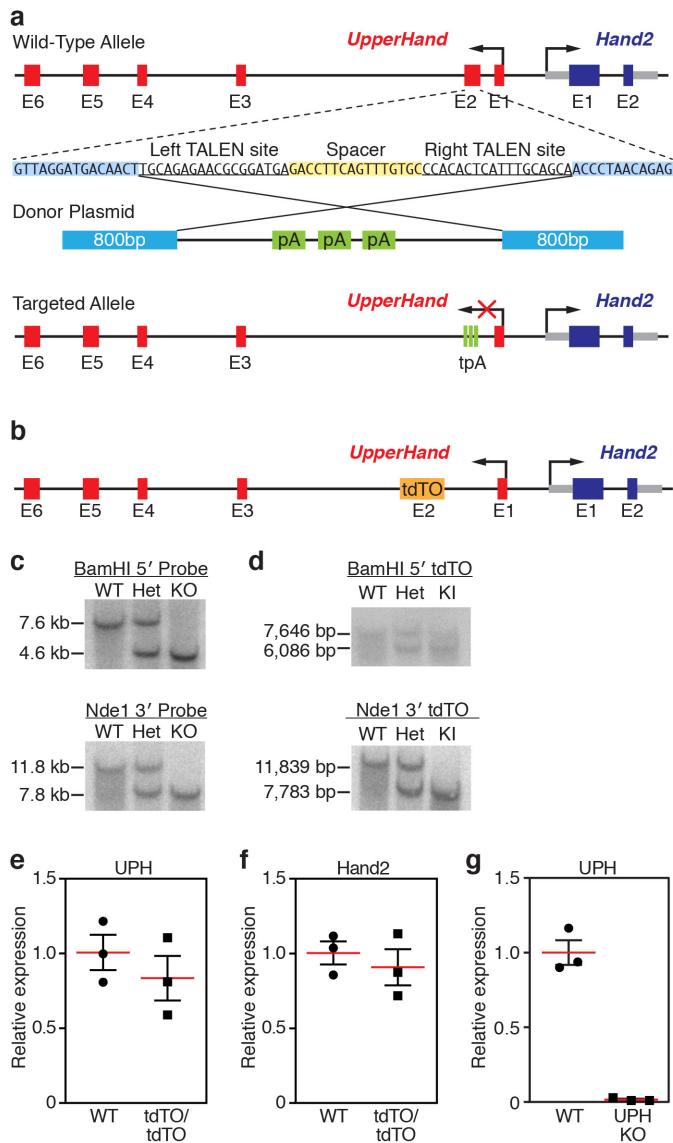
Extended Data Figure 1 | Sequence alignment of several mammalian *Uph* transcripts. **a**, Sequence alignment of several mammalian *Uph* transcripts performed using ClustalW. See Methods for source data. **b**, Diagram of the *Hand2* locus in mammals showing the genomic organization and orientation of *Uph*. The *Hand2* branchial arch enhancer (green) and cardiac enhancer (yellow) are shown. **c**, Diagram of the *Uph* transcripts expressed in the mouse heart, determined using 5' and

3' RACE using primers specific to exon 4 of *Uph*. AP1, marathon adaptor primer; 3'GSP, 3' *Uph*-specific primer from exon 4; 5'GSP, 5' *Uph*-specific primer from exon 4. **d**, Whole mount *in situ* hybridization of E10.5 mouse embryos. Expression was detected in heart, branchial arches (arrowhead), and limb bud. Scale bars, 1 mm. **e**, Northern blot analysis of total RNA from adult mouse tissues using a probe specific to the major *Uph* transcript. For gel source data, see Supplementary Fig. 1.

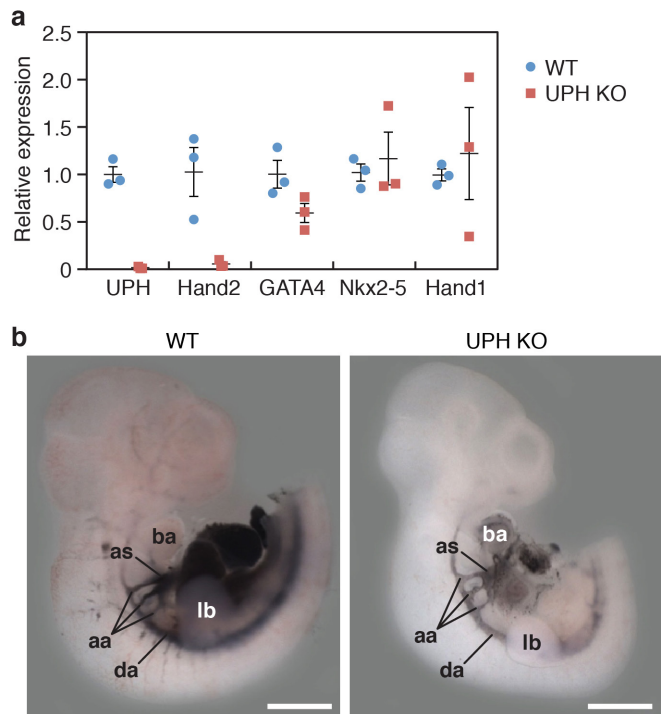


Extended Data Figure 2 | *Uph* is a cytoplasmic lncRNA. **a**, Subcellular fractionation of 18S, *Uph* and *Malat1* lncRNA in mouse neonatal cardiomyocytes ($n = 3$ biological replicates from 1 of 5 independent experiments; mean \pm s.e.m.). **b**, *In vitro* transcription and translation of a plasmid encoding the major *Uph* RNA. A plasmid encoding the

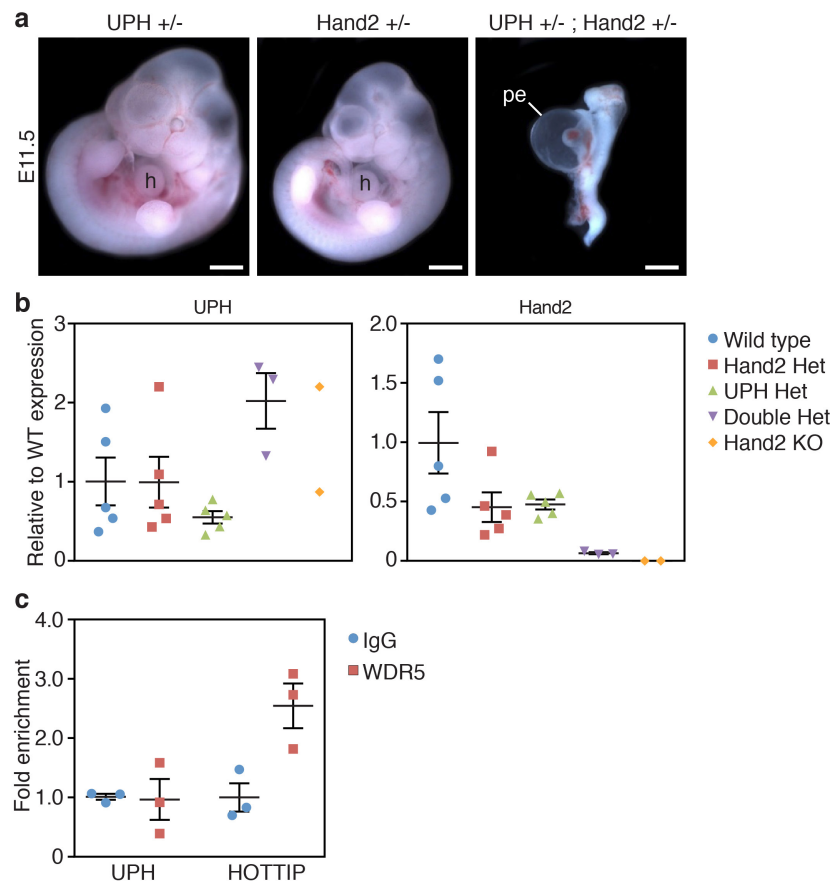
myoregulin (MLN) micropeptide was used as a positive control, and myoregulin with a frameshift mutation (MLN RNA-FS) was used as a negative control. In contrast to myoregulin, *Uph* and the negative control (MLN RNA-FS) did not produce any detectable proteins, indicating that *Uph* is a bona fide lncRNA. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 3 | Targeting strategy for insertion of transcriptional termination or heterologous sequence into exon 2 of *Uph*. **a**, *Uph* KO targeting strategy. Transcription activator-like effector nucleases (TALENs) were used to insert a triple polyadenylation (tpA) termination sequence into exon 2 (E2) of *Uph*. **b**, Using the same TALEN pair as in **a**, we introduced the coding sequence of tdTO, lacking a polyadenylation sequence, into exon 2 of the *Uph* locus. Exon 2, which includes the tdTO coding sequence, was spliced out of the mature *Uph* transcript, preventing expression of tdTO in these mice. **c**, Southern blot analysis of wild-type, heterozygous *Uph*^{+/-} (Het) and *Uph* KO genomic DNA. BamHI-digested DNA hybridized with a 5'-specific probe and NdeI-digested DNA hybridized with a 3'-specific probe. For gel source data, see Supplementary Fig. 1. **d**, Southern blot analysis of wild-type, *Uph*^{tdTO/+} heterozygous and *Uph*^{tdTO/tdTO} knock-in (KI) genomic DNA verified the correct targeting of the tdTO sequence into the *Uph* locus. DNA was digested with BamHI and hybridized with a 5'-specific probe or digested with NdeI and hybridized with a 3'-specific probe. For gel source data, see Supplementary Fig. 1. **e**, **f**, Expression of *Uph* (**e**) and *Hand2* (**f**) in wild-type and *Uph*^{tdTO/tdTO} homozygous mice at E10.0 was not changed by the insertion of the tdTO sequence into exon 2 of the *Uph* locus ($n = 3$, representative of 3 independent experiments; mean \pm s.e.m.). **g**, qPCR shows *Uph* transcripts decreased by $\sim 97\%$ in E10.5 hearts ($n = 3$ mice of each genotype from 1 of 3 independent experiments; mean \pm s.e.m.).

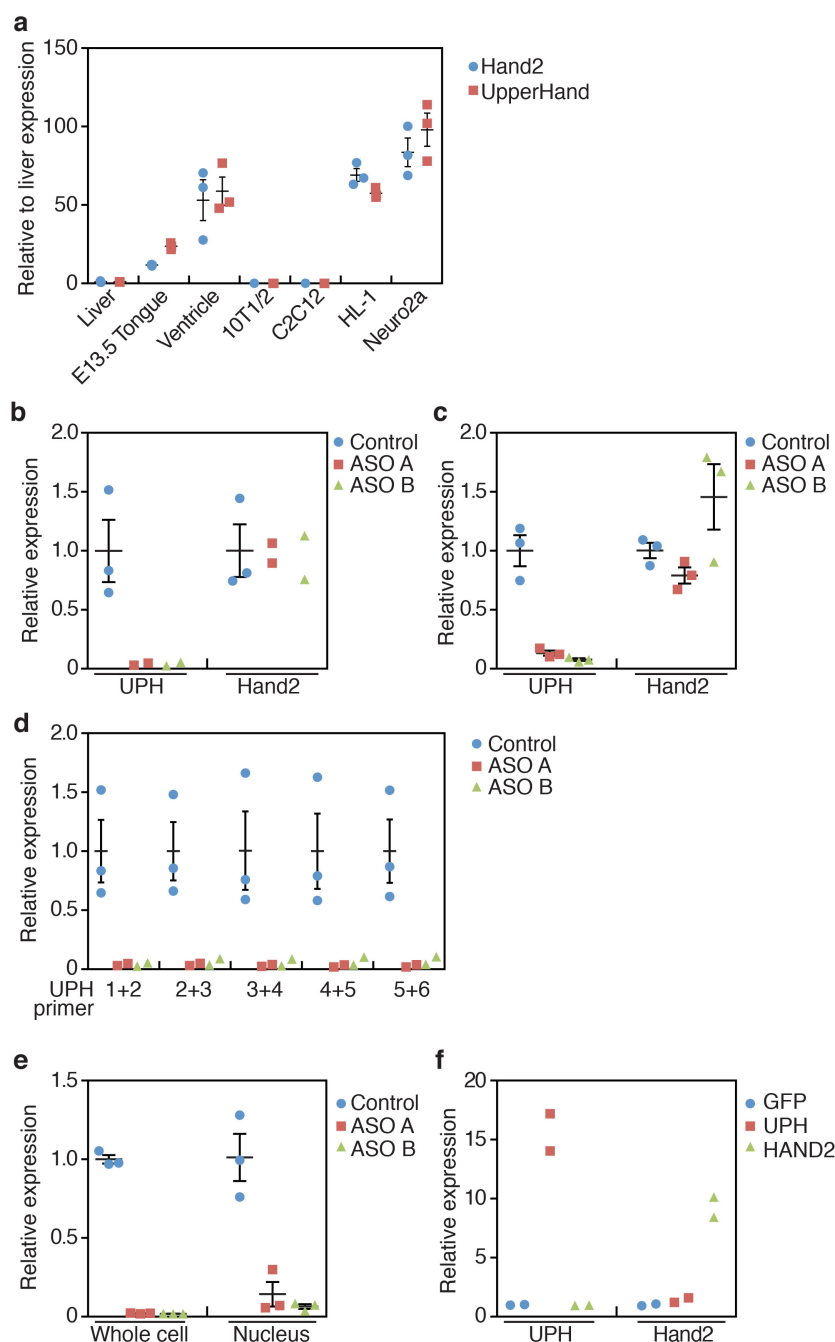


Extended Data Figure 4 | Aortic arch arteries are normal in *Uph* KO embryos. **a**, qPCR quantification of gene expression at E10.0 showed robust downregulation of *Uph* and *Hand2* expression in *Uph* KO hearts, with normal expression of other cardiac transcription factors ($n = 3$ mice of each genotype from 1 of 3 independent experiments; mean \pm s.e.m.). **b**, India ink was injected into either the left ventricle of wild-type embryos or the single ventricle of *Uph* KO embryos at E10.5, to visualize the aortic arch arteries and circulation, which appeared normal in *Uph* KO embryos. aa, aortic arch arteries; as, aortic sac; da, dorsal aorta. Scale bars, 1 mm.



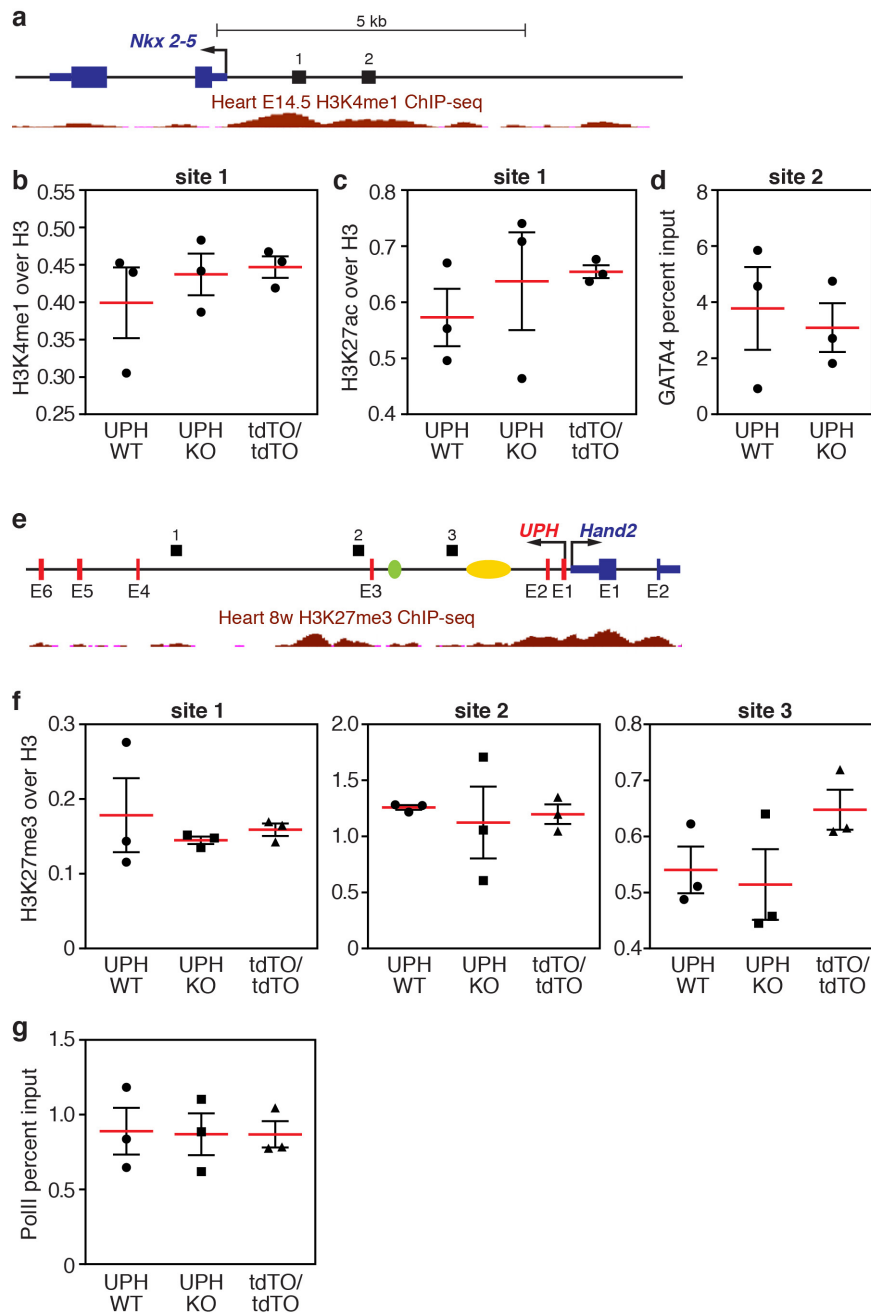
Extended Data Figure 5 | *Uph*^{+/-} *Hand2*^{+/-} compound heterozygotes recapitulate *Uph* KO phenotype. **a**, *Uph*^{+/-} *Hand2*^{+/-} double heterozygote embryos developed a single ventricle, pericardial effusion, and were severely growth restricted by E11.5. Scale bars, 1 mm. **b**, *Uph* expression is normal in double heterozygotes, whereas *Hand2* was reduced by ~90%, relative to wild-type embryos ($n = 5$ mice for wild type, *Hand2*

het and *Uph* het, $n = 3$ for double het, $n = 2$ for *Hand2* KO, from 1 of 2 independent experiments; mean \pm s.e.m.). **c**, Immunoprecipitation of RNA using either IgG or WDR5 in HL-1 cells followed by qPCR for *Uph* revealed no binding to WDR5. The WDR5-interacting lncRNA HOTTIP was used as a positive control ($n = 3$ biological replicates from 1 of 3 independent experiments; mean \pm s.e.m.).



Extended Data Figure 6 | Knockdown of mature *Uph* transcripts in HL-1 or Neuro2a cells does not affect *Hand2* expression. **a**, Expression of *Uph* (red) and *Hand2* (blue) in various tissues and cell lines. *Uph* and *Hand2* are robustly expressed in the heart, and the HL-1 and Neuro2a cell lines. *Uph* and *Hand2* transcripts are not expressed in the liver, 10T1/2 fibroblasts or skeletal muscle C2C12 myotubes ($n=3$ technical replicates; mean \pm s.e.m.). **b**, **c**, *Uph* transcripts were reduced by $\sim 90\%$ in HL-1 (**b**) and Neuro2a (**c**) cells when transfected with two independent GapmeR antisense oligonucleotide probes (ASO A and B) against *Uph*. Expression of *Hand2* was not changed ($n=2$ for ASO A and B, $n=3$ for control, from 1 of 2 independent experiments; mean \pm s.e.m.). **d**, *Uph* transcripts were similarly downregulated across each exon–exon junction, measured using

qPCR ($n=2$ for ASO A and B, $n=3$ for control, from 1 of 2 independent experiments; mean \pm s.e.m.). **e**, Fractionation of HL-1 cells transfected with control or *Uph*-specific ASOs. The nuclear fraction of antisense-oligonucleotide-treated HL-1 cells showed a similar downregulation of *Uph* transcripts ($n=3$ biological replicates from 1 of 3 independent experiments; mean \pm s.e.m.). **f**, Overexpression of enhanced green fluorescent protein (eGFP; blue), the major *Uph* RNA (red) or *Hand2* RNA (green) in HL-1 cells revealed that the mature *Uph* transcript did not alter *Hand2* expression relative to wild type, and that *Hand2* does not influence the expression of *Uph* in these cells ($n=2$ biological replicates from 1 of 2 independent experiments; mean \pm s.e.m.).



Extended Data Figure 7 | ChIP and qPCR analyses of active cardiac enhancers regulating *Nkx2-5* expression. **a**, Diagram of the *Nkx2-5* locus with numbers (1 and 2) indicating the region analysed by qPCR following ChIP. Shown in red are the ENCODE/LICR H3K4me1 active enhancer marks in the heart at E14.5. See Methods for source data. **b**, ChIP coupled with qPCR analysis of H3K4me1 marks, normalized to total histone H3, showed that the H3K4me1 marks bound to the *Nkx2-5* promoter region are unchanged between wild-type, *Uph* KO and *Uph*^{tdTO/tdTO} homozygous hearts at E10.0. **c**, ChIP coupled with qPCR analysis of H3K27ac marks, normalized to total histone H3, showed that the H3K27ac marks bound to the *Nkx2-5* promoter region are unchanged between wild-type, *Uph* KO and *Uph*^{tdTO/tdTO} homozygous hearts at E10.0. **d**, GATA4 binding to a

GATA4 site in the *Nkx2-5* promoter is unchanged between wild-type and *Uph* KO hearts at E10.0. **e**, Diagram of the *Uph-Hand2* locus, with black bars indicating the regions analysed by qPCR following ChIP. Shown in red is the ENCODE/LICR H3K27me3 track for mouse heart. See Methods for source data. The branchial arch enhancer (green) and cardiac enhancer (yellow) are shown. **f**, ChIP coupled with qPCR analysis of the polycomb repressive marker H3K27me3, normalized to total histone H3, showed no differences between genotypes at each locus (1–3) indicated in the diagram in **e**. **g**, qPCR revealed no difference in the levels of Ser2-phosphorylated RNAPII between genotypes at the *Nkx2-5* gene body. Each point is one of 3 technical replicates of 5 pooled hearts for each genotype in each ChIP experiment, from 1 of 2 independent experiments; mean ± s.e.m.

Extended Data Table 1 | Relevant sequences

SEQUENCE NAME	FORWARD	REVERSE
Full Length UPH sequence	CACTCATAACCATAAGATAATTAACCGG	TTTAAAAAATAATTTTAAATATACTATGTGCATGGTTGGATAGGT
UPH qPCR primers	CATTCTCGAGCAATTCGTCA	TGGTAGCCCATCTCCAACTC
UPH ISH probe	GATGAGACCTTCAGTTTGTGCC	ATACTATGTGCATGGTTGGATAGGT
Hand2 ISH probe	ATGAGTCTGGTGGGGGGC	TCACTGCTTGAGCTCCAGG
UPH targeting 5'arm	TATCGGAGCTCGCACCTCGGAGCTGGGAA	GATACGCGGCCCGGATCCAGTTGTCATCCTAACTTGGGTCA
UPH targeting 3'arm	TATCGGCTAGCCATATGACCCTAACAGAGATTGCGAAGA	GATACATCGATCAGGGCAGTTAGGTCTCAGC
UPH 5' WT genotyping	CTCCTCTCCGGACAAGAATC	TGCTGCAAAATGAGTGTGG
UPH 5' KO genotyping	CTCCTCTCCGGACAAGAATC	GGTTCGGATCCACTAGTTCT
UPH 3' WT genotyping	AGAGAACGCGGATGAGACCT	CCCTTGCAAAACAGAAGAAAGG
UPH 3' KO genotyping	GACCTGCAGCCCAGCTA	CCCTTGCAAAACAGAAGAAAGG
tdTO 5' WT genotyping	CTCCTCTCCGGACAAGAATC	TGCTGCAAAATGAGTGTGG
tdTO 5' KO genotyping	CTCCTCTCCGGACAAGAATC	ATGACCTCCTCGCCCTTG
tdTO 3' WT genotyping	AGAGAACGCGGATGAGACCT	CCCTTGCAAAACAGAAGAAAGG
tdTO 3' KO genotyping	GACCTGCAGCCCAGCTA	CCCTTGCAAAACAGAAGAAAGG
Southern BamHI probe	TGGTTTTCTTGTGCTTGCTG	CTGACTGGGTCTTGAGCAT
Southern NdeI probe	TCCTGGGAAGGCACTATGTG	ACCTTCTCCTGCCCTTCATT
Hand1 ISH probe	CCATCATCACCCTCACACC	GCGCCCTTTAATCCTCTTCT
Nkx2-5 ISH probe	TATGGGTACAACGCCTACCC	GTGTGGAATCCGTCGAAAGT
ASO negative control	AACACGTCTATACGC	
ASO UPH A	GCTAGTTAAGCAGGA	
ASO UPH B	ATTCAATTTAGGTCAT	
5GSP1		CAGCTGTATGGGCTCAGGTGACTGC
3GSP1		TGGAGATGGGCTACCATTGGTGTGA
AP1	CCATCCTAATACGACTCACTATAGGGC	
Hand2 - Gata site	TTTACCCACTGGTCCCTCT	TGGACAACATGGGACAGAAA
Nkx2-5 - Gata site	CTGCAACTATCACCCGAAT	AGAAACCCCATCTGTTTCC
UPH ChIP site 1	TCACCTCCCATGTCTTTTC	GAGGAACCTGCATTGCTTTC
UPH ChIP site 2	ACCTCGGGCTTTGCATCTTA	GCTTGGAAGGTAAGCCTTT
UPH ChIP site 3	GAAACTAGCCTTGCCCTTC	GGGTGCCTAGGGAGGAATAC
Nkx2-5 ChIP - site 1	ACTGTGAAGCCCAATTCCAG	AACCAGAAATGTGGCAAGG

PI3K γ is a molecular switch that controls immune suppression

Megan M. Kaneda¹, Karen S. Messer^{1,2}, Natacha Ralainirina¹, Hongying Li^{1,2}, Christopher J. Leem¹, Sara Gorjestani¹, Gyunghwi Woo¹, Abraham V. Nguyen¹, Camila C. Figueiredo^{1,3}, Philippe Foubert¹, Michael C. Schmid¹, Melissa Pink⁴, David G. Winkler⁴, Matthew Rausch⁴, Vito J. Palombella⁴, Jeffery Kutok⁴, Karen McGovern⁴, Kelly A. Frazer^{5,6}, Xuefeng Wu⁷, Michael Karin⁷, Roman Sasik⁸, Ezra E. W. Cohen^{1,9} & Judith A. Varner^{1,9,10}

Macrophages play critical, but opposite, roles in acute and chronic inflammation and cancer^{1–5}. In response to pathogens or injury, inflammatory macrophages express cytokines that stimulate cytotoxic T cells, whereas macrophages in neoplastic and parasitic diseases express anti-inflammatory cytokines that induce immune suppression and may promote resistance to T cell checkpoint inhibitors^{1–7}. Here we show that macrophage PI 3-kinase γ controls a critical switch between immune stimulation and suppression during inflammation and cancer. PI3K γ signalling through Akt and mTOR inhibits NF κ B activation while stimulating C/EBP β activation, thereby inducing a transcriptional program that promotes immune suppression during inflammation and tumour growth. By contrast, selective inactivation of macrophage PI3K γ stimulates and prolongs NF κ B activation and inhibits C/EBP β activation, thus promoting an immunostimulatory transcriptional program that restores CD8⁺ T cell activation and cytotoxicity. PI3K γ synergizes with checkpoint inhibitor therapy to promote tumour regression and increased survival in mouse models of cancer. In addition, PI3K γ -directed, anti-inflammatory gene expression can predict survival probability in cancer patients. Our work thus demonstrates that therapeutic targeting of intracellular signalling pathways that regulate the switch between macrophage polarization states can control immune suppression in cancer and other disorders.

We investigated the association between immune responses and survival in primary tumours from human papilloma virus (HPV)⁺ ($n = 97$) and HPV[−] ($n = 423$) head and neck squamous cell carcinoma (HNSCC) cohorts from the Cancer Genome Atlas^{8,9}. High expression levels of pro-inflammatory mRNAs *IL12A*, *IL12B*, *IFNG* and *CD8A* were associated with increased survival in HPV⁺ but not HPV[−] cohorts, whereas high expression of *IL6* was negatively associated with survival (Extended Data Fig. 1a–e). HPV⁺ patients with this favourable immune expression profile ($n = 35$) had 97% survival at 3 years compared with 57% survival for patients without this profile ($n = 62$) (Fig. 1a). Similar associations were observed in lung adenocarcinoma and gastric carcinoma patients (Extended Data Fig. 1f, g). These results suggest that therapeutic approaches that stimulate pro-inflammatory gene expression might enhance cancer patient survival.

We hypothesized that macrophage signalling pathways, such as those regulated by the class IB isoform PI3K γ , might control the switch between immune stimulation and suppression in inflammation and cancer. PI3K γ is abundantly expressed in myeloid cells but not in cancer cells^{10–13} (Extended Data Fig. 1h) and promotes myeloid cell trafficking during inflammation and cancer^{12–16}. Mice lacking PI3K γ (*Pik3cg*^{−/−}) had exaggerated, macrophage-mediated pro-inflammatory responses

upon exposure to pathogenic stimuli (Fig. 1b and Extended Data Fig. 1i–k), suggesting that PI3K γ inhibits macrophage inflammatory responses and might also do so in the tumour microenvironment. Mice lacking PI3K γ and mice that were treated with PI3K γ antagonists (TG100-115 (ref. 13) or IPI-549 (ref. 17)) exhibited significantly ($P < 0.05$, one-way ANOVA with Tukey's post-hoc test) suppressed growth of implanted HPV⁺ (MEER, mouse HPV⁺ HNSCC cell line) and HPV[−] (SCCVII, mouse HPV[−] HNSCC cell line) HNSCC, lung carcinoma (LLC), and breast carcinoma (PyMT) tumours (Fig. 1c and Extended Data Fig. 1l). PI3K γ inhibition did not directly affect the growth or survival of tumour cells, which do not express the kinase^{13–16} (Extended Data Fig. 1h, m). PI3K γ inhibition suppressed long-term growth and metastasis of spontaneous breast tumours, extended survival of mice with orthotopic breast tumours and enhanced the sensitivity of tumours to the nucleoside analogue gemcitabine (Fig. 1d and Extended Data Fig. 1n–q). Although PI3K γ inhibition did not affect the accumulation of CD11b⁺–Gr1[−]–F4/80⁺ tumour-associated macrophages (TAMs) in tumours (Extended Data Fig. 2a–f), it enhanced the expression of major histocompatibility complex class II (MHCII) and pro-inflammatory cytokines and inhibited the expression of immune-suppressive factors in tumours and TAMs, indicating that PI3K γ controls the TAM switch between immune suppression and immune stimulation (Fig. 1e–g and Extended Data Figs 2g–j, 3a–f).

To determine whether PI3K γ directly regulates macrophage polarization, we analysed mRNA and protein expression in primary mouse macrophages stimulated *in vitro* with basal medium containing colony stimulating factor 1 (CSF-1) or in pro-inflammatory (IFN γ , lipopolysaccharide (LPS) and CSF-1) or anti-inflammatory (interleukin-4 (IL4) and CSF-1) conditions. Pro-inflammatory stimuli upregulated macrophage expression of innate immune proteins, cytokines and cell-surface receptors, whereas anti-inflammatory stimuli induced expression of immunosuppressive factors similar to those expressed in TAMs¹ (Extended Data Fig. 4a–c). Genes and proteins associated with immune activation, antigen presentation and T cell activation were upregulated in *Pik3cg*^{−/−} and PI3K γ -inhibitor-treated macrophages (Fig. 2a–c and Extended Data Figs 4c–f, 5a–g). By contrast, genes associated with immune suppression and chemoattraction were inhibited (Fig. 2a–c and Extended Data Figs 4c–f, 5a–g). These results confirm that PI3K γ controls a macrophage switch between immune stimulation and suppression.

To investigate how PI3K γ regulates macrophage immune responses, we analysed the DNA-binding activities of NF κ B p65–RelA and the CCAAT/enhancer binding protein C/EBP β in wild-type and *Pik3cg*-null macrophages, as NF κ B promotes expression of inflammatory

¹Moore's Cancer Center, University of California, San Diego, La Jolla, California 92093, USA. ²Division of Biostatistics and Bioinformatics; Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, California 92093, USA. ³Dep. Biologia Celular, UERJ, Rio de Janeiro, 20550-013, Brazil. ⁴Infinity Pharmaceuticals, Cambridge, Massachusetts 02139, USA. ⁵Department of Pediatrics, University of California, San Diego, La Jolla, California 92093, USA. ⁶Institute for Genomic Medicine, University of California, San Diego, La Jolla, California 92093, USA. ⁷Department of Pharmacology, University of California, San Diego, La Jolla, California 92093, USA. ⁸Center for Computational Biology and Bioinformatics, Institute for Genomic Medicine, University of California, San Diego, La Jolla, California 92093, USA. ⁹Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA. ¹⁰Department of Pathology, University of California, San Diego, La Jolla, California 92093, USA.

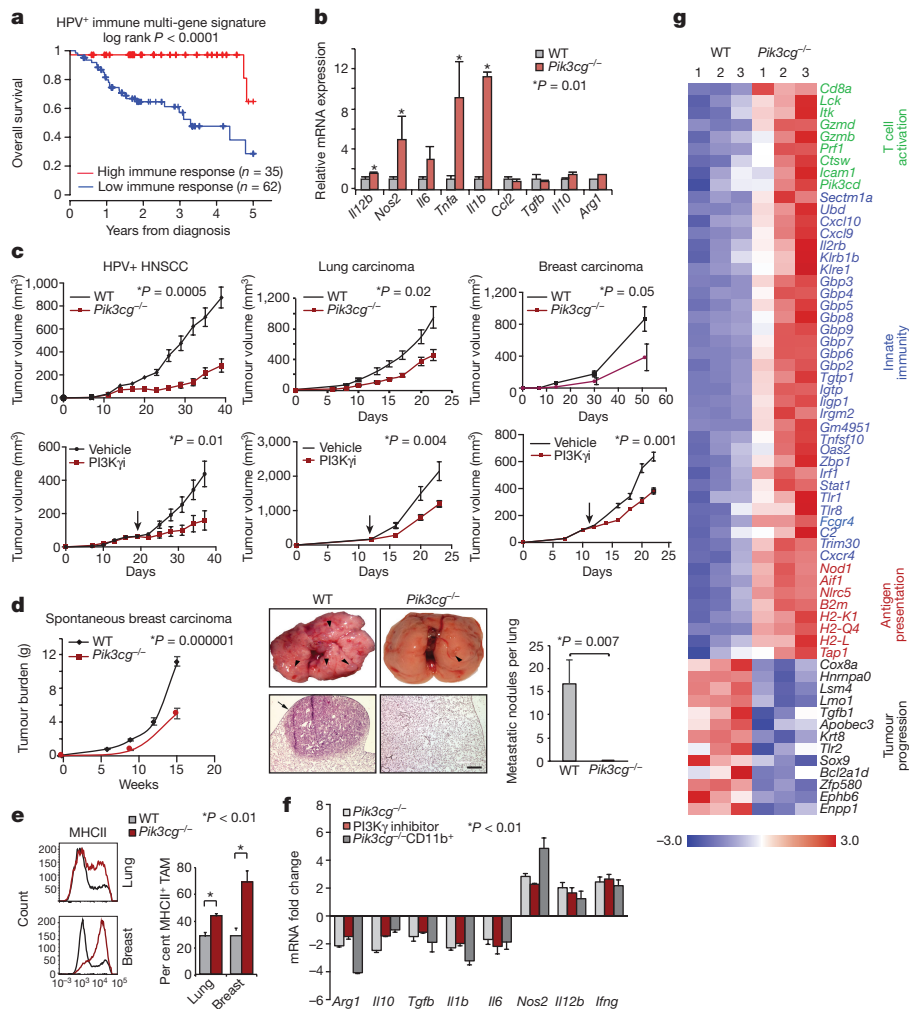


Figure 1 | PI3K γ promotes immune suppression. **a**, Multivariate immune response mRNA signature in HPV⁺ HNSCC patients. $n = 97$ biological replicates; $P = 0.0001$; log-rank test. **b**, mRNA expression of genes involved in the immune response in $Pik3cg^{-/-}$ or wild-type (WT) peritoneal macrophages. $n = 3$ biological replicates; * $P = 0.01$; t -test. **c**, Tumour volumes in wild-type, $Pik3cg^{-/-}$ and PI3K γ (PI3K γ i)-inhibitor-treated mice ($n = 15$ biological replicates) $P < 0.05$, one-sided ANOVA with Tukey's post-hoc test. Arrow shows time of initiation of daily treatment with the PI3K γ inhibitor. **d**, Tumour weight ($P = 0.000001$) and number of lung metastases ($P = 0.007$) in mice with PyMT spontaneous breast carcinoma (scale bar, 200 μ m) in wild-type ($n = 21$ biological

replicates) and $Pik3cg^{-/-}$ ($n = 8$ biological replicates) backgrounds (t -test). **e**, MHCII expression in wild-type or $Pik3cg^{-/-}$ TAMs. $n = 3$ biological replicates; $P = 0.01$ for both lung and breast tissue, as analysed by t -test. **f**, Fold change of mRNA expression in tumours and tumour-derived CD11b⁺ cells from $Pik3cg^{-/-}$ and PI3K γ -inhibitor-treated mice. $n = 5$ biological replicates; $P < 0.01$; t -test. **g**, Heat map of median-centred mRNA expression of genes involved in the immune response in tumours from wild-type and $Pik3cg^{-/-}$ mice. $n = 3$ biological replicates; $P = 0.01$; t -test. All experiments were performed two or more times. **b**, **d**, **e**, Data are shown as mean \pm s.e.m.

cytokines¹⁸, whereas C/EBP β promotes expression of the immunosuppressive factor arginase 1 (Arg1, refs 19,20). PI3K γ ablation rapidly and sustainably stimulated RelA DNA-binding activity in macrophages (Fig. 2d and Extended Data Fig. 5h). By contrast, PI3K γ ablation suppressed DNA-binding activity of C/EBP β (Fig. 2e). Consistent with these findings, PI3K γ inhibition stimulated and sustained p65–RelA phosphorylation and simultaneously inhibited C/EBP β and Akt phosphorylation (Fig. 2f, g and Extended Data Fig. 6a, b).

Subsequently, we examined the effect of PI3K γ inhibition on the stability and phosphorylation of proteins that activate NF κ B, including the TLR4-associated adaptor protein IRAK-1 and inhibitory κ B kinase β , IKK β , which promotes degradation of I κ B α , with subsequent release of NF κ B from an inhibitory I κ B–NF κ B complex¹⁸. PI3K γ deletion enhanced phosphorylation of IKK β and TBK1 and degradation of IRAK-1 and I κ B α in LPS-stimulated $Pik3cg^{-/-}$ macrophages (Fig. 2h). As an IKK β inhibitor suppressed the inflammatory phenotype observed in $Pik3cg^{-/-}$ macrophages (Fig. 2i), these results indicated that PI3K γ was both a feedback inhibitor of the TLR4–NF κ B activation pathway and a promoter of IL4 and C/EBP β signalling.

C/EBP β has previously been linked with tumour immune suppression through its control of Arg1 expression^{19,20}. Expression of constitutively activated PI3K γ ($Pik3cg^{CAAX}$)¹² was sufficient to induce Arg1 expression in a manner that was inhibited by *Cebpb*, *Mtor*, or *Rps6kb1* (also known as *S6ka*) knockdown (Extended Data Fig. 6c, d). Both *Cebpb* knockdown and inhibition of S6K α or mTOR by inhibitors suppressed expression of immune suppressive factors and stimulated expression of pro-inflammatory cytokines (Extended Data Fig. 6e–g). These results show that PI3K γ promotes immune suppression by activating mTOR–S6K α –C/EBP β and inhibiting NF κ B, thereby controlling a switch that regulates the balance between immune suppression and stimulation.

Since PI3K γ blockade stimulated pro-inflammatory responses in macrophages, we investigated whether PI3K γ blockade in macrophages promoted adaptive immunity. TAMs were isolated from wild-type and $Pik3cg^{-/-}$ tumours, mixed with tumour cells and adoptively transferred into new wild-type or $Pik3cg^{-/-}$ recipient mice (Fig. 3a). Tumour growth was significantly inhibited ($P = 0.01$) in tumours containing $Pik3cg^{-/-}$ TAMs but not wild-type TAMs (Fig. 3b). The number of

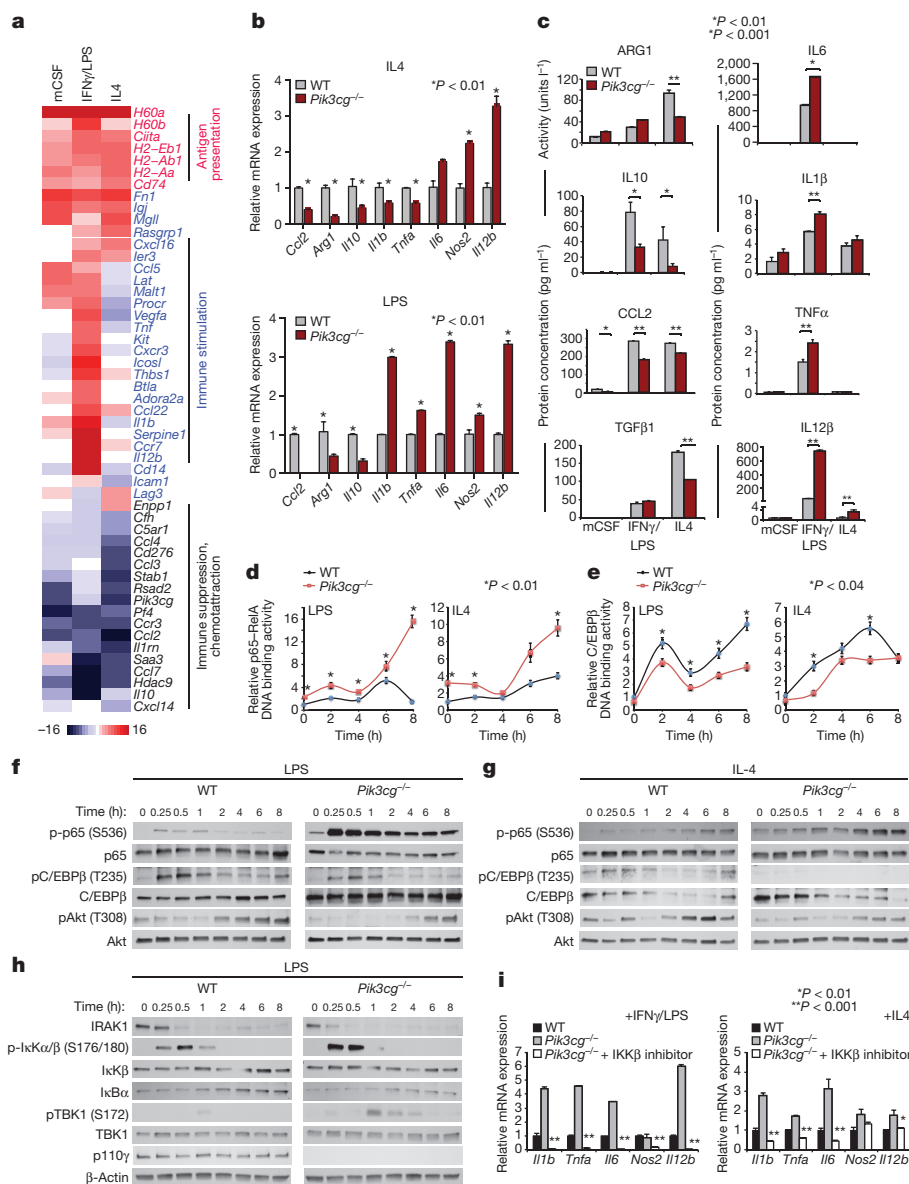


Figure 2 | PI3K γ regulates NF κ B and C/EBP β during macrophage polarization. **a**, Heat map of mean fold-change in gene expression of *Pik3cg*^{-/-} versus wild-type macrophages stimulated in mCSF, IFN γ /LPS or IL4 conditions ($n = 3$ biological replicates) for genes that are significantly differently expressed ($P = 0.00001$) in these two genotypes. **b**, mRNA expression of indicated genes in wild-type or *Pik3cg*^{-/-} mice. $n = 3$ biological replicates; $*P = 0.01$; t -test. **c**, Expression levels of indicated proteins in wild-type and *Pik3cg*^{-/-} mice. Data are mean \pm s.e.m.; $n = 4$ biological replicates; $*P = 0.01$ or $**P = 0.001$; t -test.

d, e, p65 RelA (d) and C/EBP β (e) DNA-binding activity in wild-type and *Pik3cg*^{-/-} macrophages. $n = 4$ biological replicates; $*P = 0.01$ (d) and $*P = 0.04$ (e), t -test. **f, g**, Immunoblotting of pRelA and RelA, pC/EBP β and C/EBP β , and pAkt and Akt in LPS- (f) and IL4-stimulated (g) wild-type and *Pik3cg*^{-/-} macrophages. **h**, Immunoblotting of IRAK1, pIKK β and IKK β , I κ B α , pTBK1 and TBK1, PI3K γ and actin in wild-type and *Pik3cg*^{-/-} macrophages. **i**, mRNA expression in IKK β -inhibitor-treated macrophages. $n = 3$ biological replicates; $*P = 0.01$ or $**P = 0.001$; t -test. All experiments were performed two or more times. Data are mean \pm s.e.m.

CD8⁺ T cells was significantly increased ($P = 0.0001$) in tumours with *Pik3cg*^{-/-} but not wild-type macrophages (Fig. 3c and Extended Data Fig. 6h), indicating that PI3K γ signalling in TAMs inhibits CD8⁺ T cell recruitment to tumours. To determine whether macrophage-derived cytokines controlled tumour growth, we implanted tumour cells mixed with *in vitro* cultured macrophages or in conditioned medium from cultured macrophages into wild-type mice. Tumour growth was enhanced by IL4-stimulated wild-type macrophages and conditioned medium from wild-type macrophages, but inhibited by IL4-stimulated *Pik3cg*^{-/-} macrophages and conditioned medium from *Pik3cg*^{-/-} or PI3K γ -inhibitor-treated macrophages and by all LPS-stimulated macrophages and conditioned medium from LPS-stimulated macrophages (Fig. 3d, e). To determine which macrophage-derived immune factors affect tumour growth *in vivo*, we treated wild-type and

Pik3cg^{-/-} TAMs *ex vivo* with inhibitors of mTOR, Arginase1, IKK β , IL12 or nitric oxide synthase 2 (NOS2) before mixing with tumour cells and implanting in mice (Extended Data Fig. 6i). Blockade of mTOR or Arginase1 in wild-type macrophages suppressed tumour growth, while inhibition of NOS2, IL12 or IKK β in *Pik3cg*^{-/-} macrophages stimulated tumour growth. These results indicate that PI3K γ -mTOR mediated immune suppression promotes tumour growth and that PI3K γ inhibition reverses these effects by shifting macrophages towards NF κ B-dependent pro-inflammatory polarization.

To investigate further whether macrophage PI3K γ controlled tumour growth, mice bearing pre-established tumours were treated with PI3K γ inhibitors in combination with clodronate liposomes, which deplete macrophages from tissues²¹. Separate PI3K γ inhibitor and clodronate liposome treatment each partially inhibited tumour

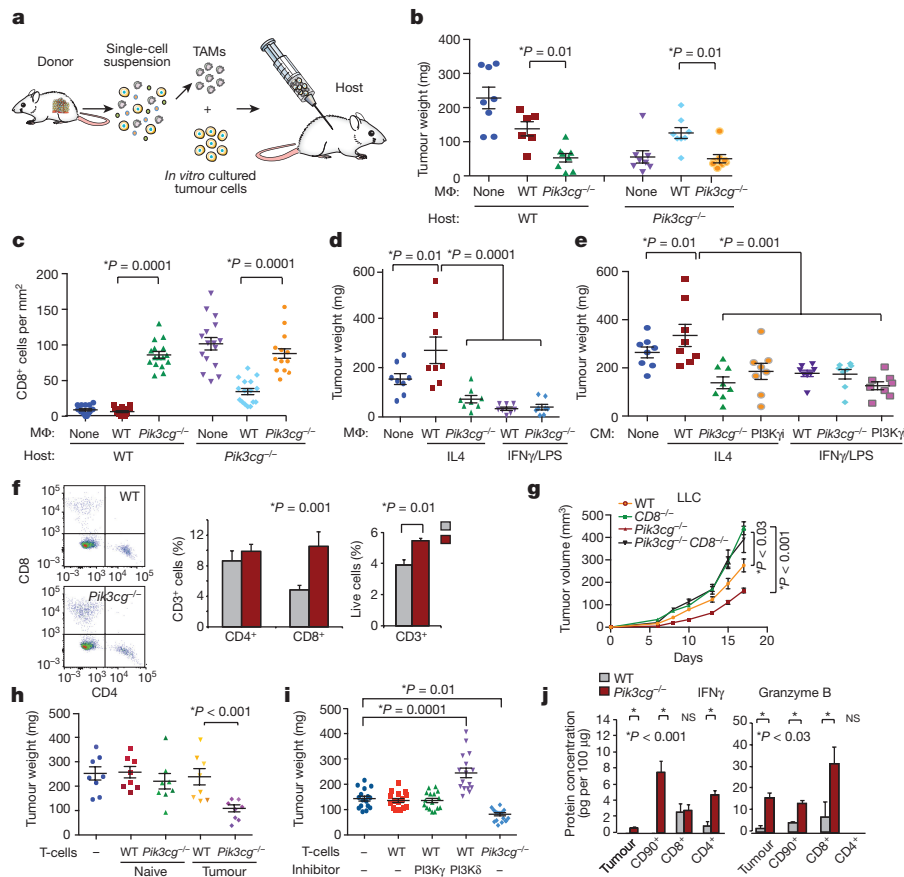


Figure 3 | Macrophage PI3K γ suppresses T cell activation. **a**, Adoptive transfer method. **b**, Weights of tumours resulting from tumour cells implanted together with *Pik3cg*^{-/-}, wild-type or no TAMs. MΦ, macrophages. *n* = 8 biological replicates; *P* = 0.01; one-sided ANOVA with Tukey's post-hoc test. **c**, CD8⁺ T cells per mm² from **b**. *n* = 8 biological replicates with 2 technical replicates each; *P* = 0.0001; one-sided ANOVA with Tukey's post-hoc test. **d**, **e**, Weights of tumours resulting from tumour cells implanted together with *in vitro* cultured macrophages (**d**, *n* = 8 biological replicates; *P* = 0.01 and *P* = 0.0001, as indicated) or with macrophage-conditioned medium (**e**, *n* = 8 biological replicates; *P* = 0.01 and *P* = 0.0001, as indicated). **f**, Percentage of CD4⁺ and CD8⁺ T cells (out of the total number of CD3⁺ cells; **P* = 0.001) and percentage of CD3⁺ T cells

(out of the total number of live cells; ***P* = 0.01; *t*-test) in wild-type and *Pik3cg*^{-/-} tumours (*n* = 3 biological replicates). **g**, Tumour volumes in *Pik3cg*^{-/-} and/or *Cd8*^{-/-} mice. *n* = 6 biological replicates; *P* < 0.03, *P* < 0.001, as indicated; one-sided ANOVA with Tukey's post-hoc test. **h**, **i**, Weights of tumours implanted with naive or tumour-derived T cells (**h**, *n* = 8 biological replicates; *P* < 0.001; *t*-test) or inhibitor-treated T cells (**i**, *n* = 16 biological replicates; **P* < 0.01; ***P* < 0.0001; one-sided ANOVA with Tukey's post-hoc test). **j**, Protein concentrations of IFN γ (**P* = 0.001; *t*-test) and granzyme B (*P* < 0.03, *t*-test) in tumours and T cells from wild-type and *Pik3cg*^{-/-} mice (*n* = 3 biological replicates; NS, not significant). All data are shown as mean \pm s.e.m. and all experiments were performed two or more times.

growth and stimulated T cell recruitment, but the combination had no additive effects, confirming that PI3K γ expression in macrophages, rather than in other cell types, promotes tumour growth (Extended Data Fig. 7a–d). Similar results were observed when CSF1R inhibition²² and PI3K γ inhibition were combined (Extended Data Fig. 7e, f).

PI3K γ blockade stimulated T cell recruitment into tumours, as total and CD8⁺ T cell content increased in tumours from *Pik3cg*^{-/-} mice without significantly altering systemic T cell content (Fig. 3f and Extended Data Fig. 7g–i). PI3K γ inhibition did not suppress tumour growth in CD8-null or antibody-depleted mice, suggesting that PI3K γ inhibition blocked tumour growth by recruiting and/or activating CD8⁺ T cells (Fig. 3g and Extended Data Fig. 7j–k). When T cells were isolated from tumour-bearing or naive mice, mixed with tumour cells and implanted in mice, only T cells from *Pik3cg*^{-/-} tumour-bearing mice suppressed tumour growth (Fig. 3h). However, PI3K γ inhibition did not directly activate T cells, as neither PI3K γ deletion nor treatment of T cells with PI3K γ inhibitors *ex vivo* affected T cell proliferation or activation; in contrast, PI3K δ inhibition suppressed T cell activation *in vitro* and promoted tumour growth *in vivo* (Fig. 3i and Extended Data Figs 7l, m, 8a, b). PI3K γ inhibition promoted T cell-mediated cytotoxicity, as T cells isolated from *Pik3cg*^{-/-} or PI3K γ -inhibitor-treated tumours stimulated tumour-cell

cytotoxicity (Extended Data Fig. 8c–g). T cells from *Pik3cg*^{-/-} or PI3K γ -inhibitor-treated mice expressed significantly more IFN γ and granzyme B (*P* = 0.001 and *P* < 0.03, respectively) and significantly less TGF β 1 and IL10 protein (*P* = 0.008 and *P* = 0.03, respectively) and mRNA (*P* < 0.02) than T cells from wild-type mice (Fig. 3j and Extended Data Fig. 8h–l). Together, these results indicate that PI3K γ inhibition in macrophages indirectly promotes both Th1 and cytotoxic adaptive immune responses.

To investigate whether PI3K γ inhibition interacts with other immune therapies, we combined PI3K γ and the checkpoint inhibitor anti-PD-1 in mouse tumour models. PD-L1, but not PD-L2, was expressed in macrophages *in vitro* and *in vivo* (Extended Data Fig. 9a, b). PI3K γ inhibition synergized with anti-PD-1 to suppress the growth of HPV⁺ HNSCC tumours in *Pik3cg*^{-/-} or inhibitor-treated male or female mice, inducing tumour regression in 86% of male and 90–100% of female mice, as well as continuous survival to date in 60% of male mice and 90–100% of female mice (Fig. 4a–c and Extended Data Fig. 9c, d). Notably, PI3K γ inhibition also synergized with anti-PD-1 to reduce tumour growth, extending survival and inducing tumour regression in 30% of mice bearing HPV⁻ HNSCC (SCCVII) tumours (Fig. 4d–f and Extended Data Fig. 9e). The combination of PI3K γ and anti-PD-1 inhibitors activated T cell memory, as 100% of mice that

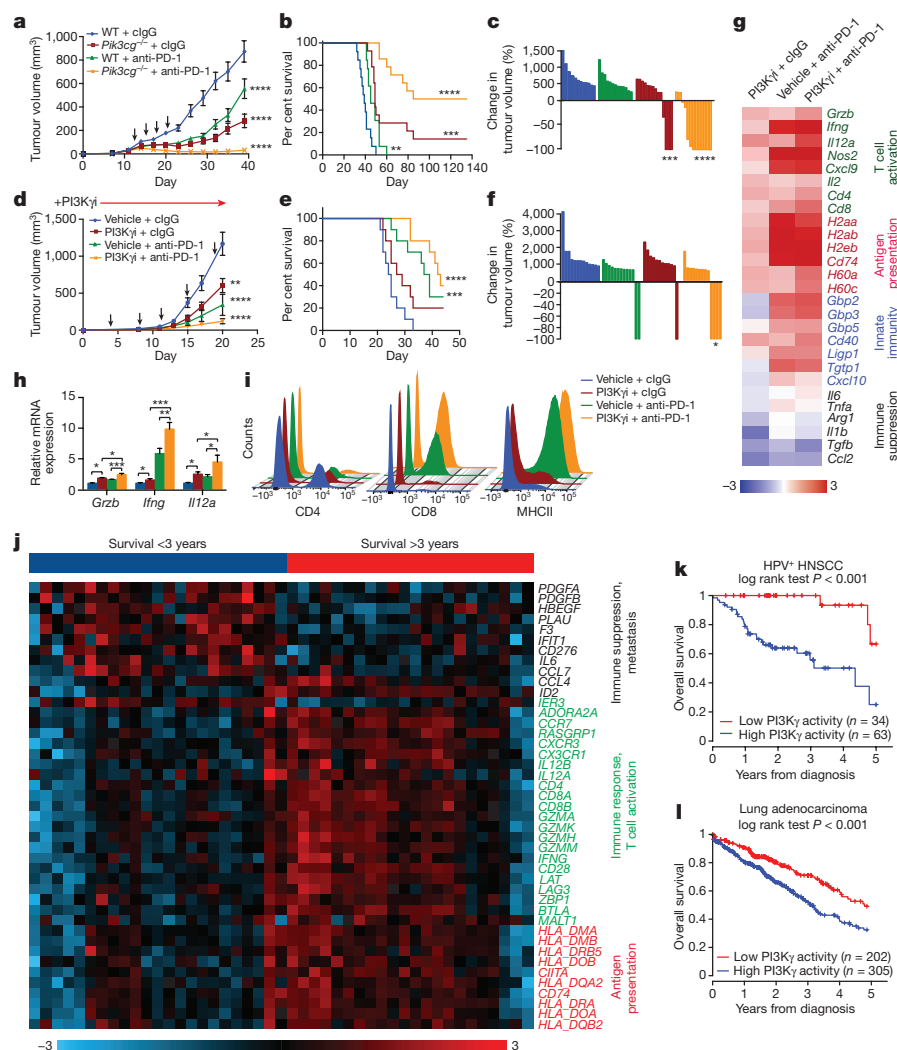


Figure 4 | PI3K γ inhibition synergizes with anti-PD-1. **a, d,** Tumour volumes in anti-PD-1 (black arrows) treated wild-type or *Pik3cg*^{-/-} mice with HNSCC HPV⁺ tumours (**a**, *n* = 13 biological replicates) and PI3K γ -inhibitor (TG100-115)-treated mice with HPV⁻ HNSCC tumours (**d**, *n* = 13 biological replicates). *****P* = 0.0001; ***P* = 0.01; one-sided ANOVA with Tukey's post-hoc test. **b, e,** Per cent survival of mice in **a** and **d**. **c, f,** Mean change in tumour volumes from individual mice from **a** and **d**. **g, h,** Median-centred heat map of whole tumour mRNA expression (**g**) and mRNA expression of select genes (**h**) from whole tumours of mice

from **d**. *n* = 3 biological replicates; **P* = 0.01; ***P* = 0.001; ****P* = 0.0001; one-sided ANOVA with Tukey's post-hoc test. **i,** Representative immune cell flow-cytometry profiles from mice from **d**. **j,** Heat map of PI3K γ -regulated mRNA expression in HPV⁺ HNSCC patients (*n* = 45 patients, *P* < 0.05; log-rank test). **k, l,** Multivariate PI3K γ -regulated immune signature in HPV⁺ HNSCC patients (**k**, *n* = 97; *P* < 0.001; log-rank test) and lung adenocarcinoma patients (**l**, *n* = 507; *P* < 0.001; log-rank test). All experiments were performed two or more times. **a, d, h,** Data are shown as mean \pm s.e.m.

had previously cleared HPV⁺ tumours efficiently suppressed tumour growth when re-challenged with HPV⁺ tumour cells and remained cancer-free (Extended Data Fig. 9f). PI3K γ and PD-1 inhibitors each stimulated immune response gene expression and inhibited immune suppressive gene expression, MHCII expression in TAMs and CD8⁺ T cell recruitment to tumours; and the combination of therapies further elevated these parameters (Fig. 4g–i and Extended Data Fig. 9g). These studies showed that PI3K γ inhibition can synergize with T-cell-targeted therapy to promote anti-tumour immune responses that induce sustained tumour regression in mouse models of cancer.

PI3K γ -regulated immune responses might also affect outcome in cancer patients. We identified 43 PI3K γ -regulated genes that are significantly (*P* < 0.05) associated with survival in the Cancer Genome Atlas HPV⁺ HNSCC patients (Fig. 4j). HPV⁺ HNSCC patients (*n* = 34) with a low PI3K γ -activity profile showed 100% survival at 3 years, compared to 56% survival for the remaining 63 patients (Fig. 4k). In HPV⁻ HNSCC patients, 39 of these genes were significantly (*P* < 0.05) shifted in the direction of high PI3K γ activity, consistent with a pattern of pervasive immune suppression and reduced survival

in HPV⁻ disease (Extended Data Fig. 10a). In lung adenocarcinoma patients, 18 genes predicted survival; 202 patients with a low PI3K γ -activity profile had 73% survival at 3 years, compared to 55% survival for 305 patients with a high PI3K γ -activity profile (Fig. 4l). These results suggest that a PI3K γ -regulated immune suppression signature is associated with survival in cancer patients and that PI3K γ inhibitors might provide clinical benefits for cancer patients.

Here we have shown that PI3K γ regulates innate immunity during inflammation and cancer (Extended Data Fig. 10b, c). Prior studies have implicated PI3Ks in the regulation of pro-inflammatory immune responses in macrophages, as pan-PI3K inhibitors and null mutations in the PI3K γ effectors PDK1, Akt1 and TSC enhanced pro-inflammatory NF κ B-dependent transcription in macrophages^{23–25}, while inhibition of PTEN and SHIP, which oppose PI3K function, promotes immune suppression^{26,27}. As macrophage reprogramming can enhance the activity of checkpoint inhibitors in cancer^{5,14,22,28}, our studies indicate that inhibitory targeting of macrophage signalling pathways may provide novel approaches to improve the long-term survival of cancer patients.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 March; accepted 12 September 2016.

Published online 19 September 2016.

- Sica, A. & Mantovani, A. Macrophage plasticity and polarization: *in vivo* veritas. *J. Clin. Invest.* **122**, 787–795 (2012).
- Schmid, M. C. & Varner, J. A. Myeloid cells in tumor inflammation. *Vasc. Cell* **4**, 14 (2012).
- Wynn, T. A., Chawla, A. & Pollard, J. W. Macrophage biology in development, homeostasis and disease. *Nature* **496**, 445–455 (2013).
- Tabas, I. & Glass, C. K. Anti-inflammatory therapy in chronic disease: challenges and opportunities. *Science* **339**, 166–172 (2013).
- Ruffell, B. & Coussens, L. M. Macrophages and therapeutic resistance in cancer. *Cancer Cell* **27**, 462–472 (2015).
- Sharma, P. & Allison, J. P. The future of immune checkpoint therapy. *Science* **348**, 56–61 (2015).
- Topalian, S. L., Drake, C. G. & Pardoll, D. M. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* **27**, 450–461 (2015).
- The Cancer Genome Atlas mRNA expression data for head and neck squamous carcinoma was downloaded from the National Cancer Institute's Genomic Data Commons Data Portal, The Cancer Genome Atlas: Head and Neck Squamous Cell Carcinoma, <https://gdc-portal.nci.nih.gov/projects/TCGA-HNSC> (accessed on 12/23/2015).
- The Cancer Genome Atlas mRNA expression data for lung adenocarcinoma was downloaded from the National Cancer Institute's Genomic Data Commons Data Portal, The Cancer Genome Atlas: Lung Adenocarcinoma Cell Carcinoma mRNA database, <https://gdc-portal.nci.nih.gov/projects/TCGA-LUAD> (accessed 04/14/2016). The National Cancer Institute's (NCI's) Genomic Data Commons (GDC) is a genomic and transcriptomic data sharing platform.
- Martini, M., De Santis, M. C., Braccini, L., Gulluni, F. & Hirsch, E. PI3K/AKT signaling pathway and cancer: an updated review. *Ann. Med.* **46**, 372–383 (2014).
- Vanhaesebroeck, B., Stephens, L. & Hawkins, P. PI3K signalling: the path to discovery and understanding. *Nat. Rev. Mol. Cell Biol.* **13**, 195–203 (2012).
- Martin, E. L. *et al.* Phosphoinositide-3 kinase gamma activity contributes to sepsis and organ damage by altering neutrophil recruitment. *Am. J. Respir. Crit. Care Med.* **182**, 762–773 (2010).
- Schmid, M. C. *et al.* Receptor tyrosine kinases and TLR/IL1Rs unexpectedly activate myeloid cell PI3K γ , a single convergent point promoting tumor inflammation and progression. *Cancer Cell* **19**, 715–727 (2011).
- Gunderson, A. J. *et al.* Bruton tyrosine kinase-dependent immune cell cross-talk drives pancreas cancer. *Cancer Discov.* **6**, 270–285 (2016).
- Schmid, M. C. *et al.* PI3-kinase γ promotes Rap1a-mediated activation of myeloid cell integrin $\alpha 4 \beta 1$, leading to tumor inflammation and growth. *PLoS One* **8**, e60226 (2013).
- Kaneda, M. M. *et al.* Macrophage PI3K γ drives pancreatic ductal adenocarcinoma progression. *Cancer Discov.* **6**, 870–885 (2016).
- Evans, C. A. *et al.* Discovery of a selective phosphoinositide-3-kinase (PI3K) γ inhibitor (IPI-549) as an immuno-oncology clinical candidate. *ACS Med. Chem. Lett.* **7**, 862–867 (2016).
- Ben-Neriah, Y. & Karin, M. Inflammation meets cancer, with NF- κ B as the matchmaker. *Nat. Immunol.* **12**, 715–723 (2011).
- Poli, V. The role of C/EBP isoforms in the control of inflammatory and native immunity functions. *J. Biol. Chem.* **273**, 29279–29282 (1998).
- Gray, M. J., Poljakovic, M., Kepka-Lenhart, D. & Morris, S. M. Jr. Induction of arginase I transcription by IL-4 requires a composite DNA response element for STAT6 and C/EBP β . *Gene* **353**, 98–106 (2005).
- van Rooijen, N., Kors, N., ter Hart, H. & Claassen, E. *In vitro* and *in vivo* elimination of macrophage tumor cells using liposome-encapsulated dichloromethylene diphosphonate. *Virchows Arch. B Cell Pathol. Incl. Mol. Pathol.* **54**, 241–245 (1988).
- Pyonteck, S. M. *et al.* CSF-1R inhibition alters macrophage polarization and blocks glioma progression. *Nat. Med.* **19**, 1264–1272 (2013).
- Chaurasia, B. *et al.* Phosphoinositide-dependent kinase 1 provides negative feedback inhibition to Toll-like receptor-mediated NF- κ B activation in macrophages. *Mol. Cell. Biol.* **30**, 4354–4366 (2010).
- Arranz, A. *et al.* Akt1 and Akt2 protein kinases differentially contribute to macrophage polarization. *Proc. Natl Acad. Sci. USA* **109**, 9517–9522 (2012).
- Byles, V. *et al.* The TSC-mTOR pathway regulates macrophage polarization. *Nat. Commun.* **4**, 2834 (2013).
- Yue, S. *et al.* Myeloid PTEN deficiency protects livers from ischemia reperfusion injury by facilitating M2 macrophage differentiation. *J. Immunol.* **192**, 5343–5353 (2014).
- Rauh, M. J. *et al.* SHIP represses the generation of alternatively activated macrophages. *Immunity* **23**, 361–374 (2005).
- Baer, C. *et al.* Suppression of microRNA activity amplifies IFN- γ -induced macrophage activation and promotes anti-tumour immunity. *Nat. Cell Biol.* **18**, 790–802 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by NIH grants R01CA126820 (J.A.V.), T32HL098062 (M.M.K.), T32CA009523 (S.G.) and T32CA121938 (S.G.), the CAPES Foundation and Ministry of Education of Brazil (C.F.) and by Ralph and Fernanda Whitworth and the Immunotherapy Foundation (J.A.V. and E.E.C.). The authors thank J. Lee and S. Schoenberger for HPV⁺ MEER HNSCC and SSCVII cells.

Author Contributions TCGA analysis was performed by H.L. and K.S.M., RNA sequencing by K.A.F., M.M.K., S.G. and R.S., flow cytometry by M.M.K. and N.R., *in vitro* studies by M.M.K., N.R., S.G., G.W., C.C.F., A.V.N. and M.C.S., and animal studies by M.M.K., N.R., C.L. and P.F. M.P., V.J.P., J.K., K.M., M.R. and D.G.W. provided IPI-549 and carried out experiments for Fig. 1c, Extended Data Fig. 8a–b. ML120B was contributed by X.W. and M.K. The project was directed by E.E.W.C., K.S.M. and J.A.V. The manuscript was written by J.A.V. and M.M.K.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.V. (jvarner@ucsd.edu).

Reviewer Information *Nature* thanks F. Balkwill, M. de Palma and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Immune-related gene expression signature analysis in the Cancer Genome Atlas (TCGA) data. We analysed TCGA data for association between mRNA expression level of 16 candidate immune-related genes (*ARG1*, *IL10*, *FOXP3*, *CD68*, *IL12A*, *IL12B*, *IFNG*, *CD8A*, *CD4*, *ITGAM* (also known as *CD11B*), *CD14*, *TNF*, *IL1A*, *IL1B*, *IL6* and *CCL5*) and 5 year overall survival. Illumina HiSeq RNAseqV2 mRNA expression and clinical data for 520 head and neck squamous cell carcinoma samples were downloaded from the TCGA data portal. Median follow-up from diagnosis was 1.8 years with a range of 0.01 years to 17.6 years. Follow-up time was truncated at 5-years for analysis and 200 deaths occurred in this period. For each of the 16 candidate immune response genes, we scored subjects as above (high) or below (low) the median expression and compared survival using a log-rank test at 5% significance. HPV⁺ patients were stratified into a favourable immune profile if they had expression above the median for the significant genes *IL12A*, *IL12B*, *IFNG*, *CD8A* and below the median for *IL6*. Kaplan–Meier curves were plotted for these two groups. Similar methods were used to examine association of these 16 genes with 720 lung adenocarcinoma and 876 gastric carcinoma samples using the publically available data from KM Plotter²⁹. In lung adenocarcinomas, 12 genes were significantly associated with survival; patients were scored as having a favourable immune profile if 7 or more of the 12 significant genes had expression in the favourable direction. In 876 gastric cancer samples, 8 genes were significantly associated with survival. Patients were scored as having a favourable immune profile if 5 out of the 8 genes had expression in the favourable direction.

PI3K γ -regulated gene expression signature analysis in TCGA data. We investigated 66 immune-related genes in four functional classes, 17 genes related to antigen presentation (HLA class I and II molecules), 24 genes surveying T cell activation, 20 innate immune response genes (*IL6*, *CCL7* and others) and 5 genes related to cancer cell signalling. These genes changed expression in response to PI3K γ inhibition for association with survival in HPV⁺ and HPV[−] TCGA HNSCC and lung adenocarcinoma cohorts. Within each cancer type, we scored subjects as above or below the median expression for each gene and compared survival using a log-rank test, using 10% false discovery rate (FDR) within each class as the significance threshold. HPV⁺ and HPV[−] HNSCC survival were investigated separately, as HPV[−] HNSCC generally has a worse prognosis. Within each cohort, patients were classified as having a favourable PI3K γ immune response profile if they had expression levels above or below the median in the direction of low PI3K γ activity for the genes identified as significant. We compared the survival experience of favourable versus less-favourable profiles of patients using Kaplan–Meier curves. Out of the 66 experimentally identified PI3K γ -regulated genes, 43 showed significant association with overall survival in the HPV⁺ cohort (FDR < 10% within each functional class). Comparison of these genes between HPV⁺ and HPV[−] cohorts showed that HPV[−] samples generally had significantly ($P < 0.05$) lower expression of 42 genes in the antigen presentation and T cell activation classes, consistent with a pattern of adaptive immune suppression, and higher expression of genes in the innate immune response and cancer cell signalling classes, which were negatively associated with survival. Only *MALT1* was not differentially expressed between the two groups ($P = 0.7$).

Mice. *Pik3cg*^{−/−} and *Pik3cg*^{−/−}, PyMT mice were generated as previously described¹³. *Cd8*^{−/−} and *Cd4*^{−/−} mice with a C57BL/6J background were purchased from the Jackson Laboratory and crossed with syngeneic *Pik3cg*^{−/−} mice. All animal experiments were performed with approval from the Institutional Animal Care and Use Committee of the University of California. Animals were euthanized before the IACUC maximum allowable tumour burden of 2 cm³ per mouse was exceeded.

Tumour studies. Wild-type or *Pik3cg*^{−/−} 6–8 week-old female or male syngeneic C57BL/6J (LLC lung, PyMT breast and MEER HPV⁺ HNSCC) or C3He/J (SCCVII HPV[−] HNSCC) mice were implanted with 10⁶ tumour cells by subcutaneous injection (LLC, MEER, SCCVII) or by orthotopic injection (PyMT) ($n = 10$ –15) and tumour growth was monitored for up to 30 days. Tumour dimensions were measured once when tumours were palpable. Tumour volumes were calculated using the equation ($l^2 \times w$)/2. In some studies, wild-type and *Pik3cg*^{−/−} mice with LLC tumours were treated with gemcitabine (150 mg kg^{−1}) or saline by intraperitoneal (i.p.) injection on day 7 and day 14 ($n = 10$). LLC were acquired from ATCC, PyMT were from L. Ellies (University of California), HPV⁺ MEER were from J. Lee (Cancer Biology Research Center, Sanford Research/USD) and SCCVII squamous carcinoma cells were from S. Schoenberger (La Jolla Institute for Allergy and Immunology). All cell lines were tested for mycoplasma and mouse pathogens and checked for authenticity against the International Cell Line Authentication Committee (ICLAC; <http://iclac.org/databases/cross-contaminations/>) list.

In some studies, mice bearing LLC, PyMT, HPV⁺ MEER or HPV[−] HNSCC tumour cells were treated once daily by oral gavage with vehicle (5% NMP and 95% PEG 400), 15 mg kg^{−1} per day of the PI3K γ inhibitor IPI-549 or by i.p.

injection with 2.5 mg kg^{−1} twice per day of TG100-115 (ref. 13) beginning on day 8 post-tumour injection and continuing daily until euthanasia. IPI-549 is an orally bioavailable PI3K γ inhibitor with a long plasma half-life and a K_D value of 0.29 nM for PI3K γ with >58-fold weaker binding affinity for the other class I PI3K isoforms¹⁷. Enzymatic and cellular assays confirmed the selectivity of IPI-549 for PI3K γ (>200-fold in enzymatic assays and >140-fold in cellular assays over other class I PI3K isoforms¹⁷). To study the effect of IPI-549 on lung tumour growth, LLC tumour cells were passaged three times in C57BL/6 albino male mice. When tumour volume reached 1,500 mm³, tumours were collected and single-cell suspensions were prepared. This tumour cell suspension was implanted subcutaneously in the hind flank of C57BL/6 albino male mice at 10⁶ cells per mouse. Prior to initiating treatment with once daily IPI-549 (15 mg kg^{−1} orally), groups were normalized on the basis of tumour volume. In some studies, wild-type- and *Pik3cg*^{−/−} tumour-bearing mice were treated with 100 μ g of anti-CD8 (clone YTS 169.4) or an isotype-control clone (LTF-2) from Bio X Cell administered by i.p. injections on day 7, 10 and 13 of tumour growth. For all tumour experiments, tumour volumes and weights were recorded at death.

Anti-PD-1 tumour studies. C57BL/6J (wild-type) or *Pik3cg*^{−/−} 6–8 week-old male or female mice (MEER HPV⁺ HNSCC) or C3He/J (SCCVII HPV[−] HNSCC) were implanted with tumour cells by subcutaneous injection (10⁶ MEER or 10⁵ SCCVII). In HPV⁺ MEER studies, wild-type and *Pik3cg*^{−/−} mice were treated with four doses of 250 μ g of anti-PD-1 antibody (clone RMP-14, Bio X Cell) or rat IgG2a isotype control (clone 2A3, Bio X Cell) every 3 days, starting when tumours became palpable on day 11 ($n = 12$ –14 mice per group). Wild-type mice bearing HPV⁺ tumours were also treated with the PI3K γ inhibitor TG100-115 (ref. 13) twice per day by i.p. injection, beginning on day 11. Tumour regressions were calculated as a percentage of the difference in tumour volume between the date treatment was initiated and the first date of death of the control group. For HPV[−] SCCVII studies, C3He/J mice were treated with PI3K γ inhibitor (2.5 mg kg^{−1} TG100-115 i.p.) beginning on day 6 post-tumour inoculation and with six doses of anti-PD-1 antibody (250 μ g clone RMP-14, Bio X Cell) or rat IgG2a isotype control (clone 2A3, Bio X Cell) every 3 days beginning on day 3 ($n = 12$ mice per group) or with a combination of the two. Alternatively, mice were treated with 5 mg kg^{−1} TG100-115 twice per day \pm anti-PD-1 (250 μ g every 3 days) beginning on day 1 (Fig. 4). Mice that completely cleared HPV⁺ MEER tumours were re-injected with HPV⁺ tumour cells contralateral to the initial tumour injection and tumour growth was monitored.

PyMT models of mammary carcinoma. The growth and metastasis of spontaneous mammary tumours in female PyMT⁺ ($n = 13$) and *Pik3cg*^{−/−}, PyMT⁺ ($n = 8$) mice was evaluated over the course of 0–15 weeks. Total tumour burden was determined by subtracting the total mammary gland mass in PyMT[−] mice from the total mammary gland mass in PyMT⁺ mice. Lung metastases were quantified macroscopically and microscopically in H&E tissue sections at week 15.

LPS-induced septic shock. Septic shock was induced in wild-type and *Pik3cg*^{−/−} mice via i.p. injection of 25 mg kg^{−1} LPS (Sigma, B5:005). Survival was monitored every 12 h and liver, bone marrow and serum were collected 24 h after LPS injection.

Macrophage depletion studies. C57BL/6J female mice were implanted with 10⁶ LLC tumour cells by subcutaneous injection. When the average tumour size was 250 mm³, mice were treated by i.p. injection with 1 mg per mouse clodronate or control liposomes (www.clodronateliposomes.com) every 4 days for 2 weeks in combination with daily administration of vehicle or IPI-549 (15 mg kg^{−1} per day orally). In other studies, 6-week-old female BALB/c mice were injected subcutaneously with 2.5×10^5 CT26 mouse colon carcinoma cells in 100 μ l phosphate buffered saline (PBS) in the right flank. Eight days later, tumour-bearing mice were arranged into four groups ($n = 15$) with an average tumour volume of 70 mm³. Oral administration of IPI-549 (15 mg kg^{−1}) or vehicle (5% NMP and 95% PEG 400) and anti-CSF-1R antibody (50 mg kg^{−1} i.p. 3 \times per week, clone AFS98, Bio X Cell) began on day 8 after tumour injection via oral gavage at a 5 ml kg^{−1} dose volume and continued daily for a total of 18 doses.

Tumour-infiltrating myeloid cell analysis. Six-week-old female BALB/c mice were injected subcutaneously with 2.5×10^5 CT26 mouse colon carcinoma cells in 100 μ l PBS in the right flank. On day 8 after tumour injection, tumour-bearing mice were grouped and treated daily with IPI-549 (15 mg kg^{−1}, orally) or vehicle (5% NMP and 95% PEG 400). In addition, mice were injected i.p. with 50 mg kg^{−1} anti-CD115 (Bio X Cell clone AFS98) or 50 mg kg^{−1} rat IgG2a isotype control (Bio X Cell clone 2A3) antibodies as described above for a total of three injections. Two days after the final injection mice were euthanized, tumours were digested in a mixture of 0.5 mg ml^{−1} collagenase IV and 150 U ml^{−1} DNase I in RPMI-1640 for 30 min at 37 °C and tumour-infiltrating myeloid cells were analysed by flow cytometry.

In vivo macrophage adoptive transfer experiments. CD11b⁺Gr1[−] cells were isolated from single-cell suspensions of LLC tumours from donor mice by

fluorescence-activated cell sorting (FACS) or serial magnetic bead isolation. Additionally, for some experiments, primary bone-marrow-derived macrophages were polarized and collected into a single-cell suspension. Purified cells were admixed 1:1 with LLC tumour cells and 5×10^5 total cells were injected subcutaneously into new host mice. Tumour dimensions were measured three times per week beginning on day 7. In antibody blocking studies, CD11b⁺Gr1[−] cells were incubated with 5 µg anti-IL12 (clone RD1-5D9) or isotype (clone LTF-2, Bio X Cell) for 30 min before the addition of tumour cells. Mice were additionally treated intradermally with 5 µg of antibody 3 and 6 days after tumour cell inoculation. In some studies, CD11b⁺Gr1[−] cells were pre-incubated with inhibitors of arginase (nor-NOHA, 50 µM, Cayman Chemical), iNOS (1400W dihydrochloride, 100 µM, Tocris), mTOR (rapamycin, 10 µM Calbiochem), or IκKβ (ML120B, 30 µM, Tocris) for 30 min before the addition of tumour cells. Inoculated mice were further treated by intradermal injection with inhibitors at 3 and 6 days after inoculation.

T cell adoptive transfer. Donor C57BL/6J (WT) or *Pik3cg*^{−/−} mice were implanted with 10⁶ LLC tumour cells by subcutaneous injection. On day 14 after tumour implantation, CD90.2⁺, CD4⁺ or CD8⁺ cells were harvested by magnetic bead isolation (Miltenyi Biotec). T cells were mixed 1:1 with viable LLC tumour cells. Cell mixtures containing 5×10^5 total cells were injected into the flanks of naive wild-type or *Pik3cg*^{−/−} mice ($n = 8–10$ per group). Tumour growth, intratumoral apoptosis and necrosis were investigated over 0–16 days. In other studies, wild-type T cells were incubated at 37°C and 5% CO₂ for 6 h with 10 or 100 nM IPI-549 (Infinity Pharmaceuticals) or Cal-101 (Selleck Chem). After 6 h, T cells were washed, admixed 1:1 with LLC tumour cells, and 10⁶ total cells were injected subcutaneously into recipient mice. Tumour growth was monitored for 14 days.

Isolation of single cells from mouse tumours. Tumours were isolated, minced in a Petri dish on ice and then enzymatically dissociated in Hanks balanced salt solution containing 0.5 mg ml^{−1} collagenase IV (Sigma), 0.1 mg ml^{−1} hyaluronidase V (Sigma), 0.6 U ml^{−1} dispase II (Roche) and 0.005 MU ml^{−1} DNase I (Sigma) at 37°C for 5–30 min. The duration of enzymatic treatment was optimized for greatest yield of live CD11b⁺ cells per tumour type. Cell suspensions were filtered through a 70-µm cell strainer. Red blood cells were solubilized with red cell lysis buffer (Pharm Lyse, BD Biosciences) and the resulting suspension was filtered through a cell strainer to produce a single-cell suspension. Cells were washed once with PBS before use in flow cytometry analysis or magnetic bead purification.

Peritoneal macrophage isolation. Thioglycollate-elicited peritoneal macrophages were collected 96 h after i.p. injection of a 3% thioglycollate solution. Cells were collected from the peritoneal cavity in 10 ml of PBS and macrophage enrichment was performed by plating cells in RPMI with 10% FBS and 1% penicillin/streptomycin for 2 h at 37°C and 5% CO₂. After 2 h, non-adherent cells were removed with three PBS washes, and cells were analysed via flow cytometry and qPCR analysis.

Flow cytometry staining and analysis. Single-cell suspensions (10⁶ cells in 100 µl total volume) were incubated with aqua live dead fixable stain (Life Technologies), FcR-blocking reagent (BD Biosciences) and fluorescently labelled antibodies and incubated at 4°C for 1 h. Primary antibodies to cell surface markers directed against F4/80 (BM8), CD45 (30-F11), CD11b (M1/70), Gr1 (RB6-8C5), CD3 (145-2C11), CD4 (GK1.5), CD8 (53-6.7), CD273 (B7-DC), CD274 (B7-H1) were from eBioscience; Ly6C (AL-21), Ly6G (1A8), CD11c (HL3), and MHC-II (AF6-120.1) from BD Pharmingen, CCR2 (475301) from R&D Systems and CD206 (MR5D3) from AbD Serotec. For intracellular staining, cells were fixed, permeabilized using transcription factor staining buffer set (eBioscience) and then incubated with fluorescently labelled antibodies to FoxP3 (FJK-16s) from eBioscience. Multicolour FACS analysis was performed on a BD Canto RUO 11 colour analyser. All data analysis was performed using the flow cytometry analysis program FloJo (Treestar).

Magnetic bead purification of myeloid cells. Single-cell preparations from bone marrow or tumours were incubated with FcR-blocking reagent (BD Biosciences) and then with 20 µl magnetic microbeads conjugated to antibodies against CD11b, Gr1, CD90.2, CD4 and CD8 (Miltenyi Biotec MACS Microbeads) per 10⁷ cells for 20 min at 4°C. Cells bound to magnetic beads were then removed from the cell suspension according to the manufacturer's instructions.

Flow cytometric sorting of cells from tumours and bone marrow. For cell sorting, single-cell suspensions were stained with aqua live dead fixable stain (Life Technologies) to exclude dead cells and anti-CD11b-APC (M1/70, eBioscience) and anti-Gr1-FITC (RB6-8C5, eBioscience) antibodies. FACS sorting was performed on a FACS Aria 11 colour high speed sorter at the Flow Cytometry Core at the UC San Diego Center for AIDS Research. Live cells were sorted into the following populations: CD11b⁺Gr1[−], CD11b⁺Gr1^{lo}, CD11b⁺Gr1^{hi} and CD11b[−]Gr1[−] cells. CD11b-positive cells were defined by increased staining over the isotype control, and Gr1 levels were defined both by comparison to the isotype control and relative staining to other populations.

Mouse macrophage differentiation and culture. Bone-marrow-derived cells were aseptically collected from 6–8 week-old female mice by flushing leg bones

of euthanized mice with PBS, 0.5% BSA, 2 mM EDTA, incubating in red cell lysis buffer (155 mM NH₄Cl, 10 mM NaHCO₃ and 0.1 mM EDTA) and centrifuging over Histopaque 1083 to purify the mononuclear cells. Approximately 5×10^7 bone-marrow-derived cells were purified by gradient centrifugation from the femurs and tibias of a single mouse. Purified mononuclear cells were cultured in RPMI + 20% serum + 50 ng ml^{−1} mCSF (PeproTech).

Human macrophage differentiation and culture. Human leukocytes from apheresis blood products were obtained from the San Diego Blood Bank. Cells were diluted in PBS, 0.5% BSA, 2 mM EDTA, incubated in red cell lysis buffer (155 mM NH₄Cl, 10 mM NaHCO₃ and 0.1 mM EDTA) and centrifuged over Histopaque 1077 to purify mononuclear cells. Approximately 10⁹ bone-marrow-derived cells were purified by gradient centrifugation from one apheresis sample. Purified mononuclear cells were cultured in RPMI + 20% serum + 50 ng ml^{−1} Human mCSF (PeproTech). Non-adherent cells were removed after 2 h by washing and adherent cells were cultured for 6 days to differentiate macrophages fully.

Macrophage polarization. Bone-marrow-derived macrophages were polarized with IFNγ (20 ng ml^{−1}, Peprotech) + LPS (100 ng ml^{−1}, Sigma) or LPS alone for 24 h, or IL4 (20 ng ml^{−1}, Peprotech) for 24–48 h. For inhibitor studies, PI3Kγ inhibitors (1 µM) (IPI-549, Infinity Pharmaceuticals and TG100-115, Targegen/Sanofi-Aventis), rapamycin (10 µM, Selleck), or ML120B (30 µM) were incubated with macrophages 1 h before the addition of polarizing stimuli. Total RNA was harvested from macrophages using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions.

RNA sequencing. Freshly isolated mouse bone marrow cells from nine wild-type and nine *Pik3cg*^{−/−} mice were pooled into three replicates sets of wild-type or *Pik3cg*^{−/−} cells and differentiated into macrophages for 6 days in RPMI + 20% FBS + 1% penicillin/streptomycin + 50 ng ml^{−1} mCSF. Each replicate set of macrophages was then treated with mCSF, IL4 or IFNγ/LPS. Macrophages were removed from dishes, and RNA was collected using Qiagen Allprep kit. In addition, RNA was harvested from day 14 (500 mm³) LLC tumours or purified CD11b⁺Gr1-F480⁺ TAMs from wild-type (C57BL/6) and *Pik3cg*^{−/−} mice. RNA was collected using the Qiagen Allprep kit. RNA libraries were prepared from 1 µg RNA per sample for sequencing using standard Illumina protocols. RNA sequencing was performed by the University of California, San Diego Institute for Genomic Medicine. mRNA profiles were generated by single read deep sequencing, in triplicate, using Illumina HiSeq2000.

Sequence analysis. Sequence analysis was performed as previously described¹⁶. Sequence files from Illumina HiSeq that passed quality filters were aligned to the mouse transcriptome (mm9 genome build) using the Bowtie2 aligner⁴. Gene-level count summaries were analysed for statistically significant changes using DESeq. Individual P values were adjusted for multiple testing by calculating Storey's q values using fdrtooltrimmer. For each gene, the q value is the smallest false discovery rate at which the gene is found significant. We analysed biological processes as defined by the Gene Ontology Consortium. Each gene ontology term defines a set of genes. The entire list of genes, sorted by the q value in ascending order, was subjected to a non-parametric variant of the gene set enrichment analysis (GSEA), in which the parametric Kolmogorov–Smirnov P value was replaced with the exact rank-order P value. We perform a Bonferroni adjustment of gene set P values for the number of gene sets tested. Heat maps of expression levels were created using in-house hierarchical clustering software that implements Ward clustering. The colours qualitatively correspond to fold changes.

Individual quantitative RT-PCR. cDNA was prepared using 1 µg RNA with the qScript cDNA Synthesis Kit (Quanta Biosciences). Sybr green-based qPCR was performed using human and mouse primers to *Arg1*, *Ifng*, *Il10*, *Il21p40*, *Il1b*, *Il6*, *Ccl2*, *Vegfa*, *Gapdh*, *Nos2*, *Tgfb1*, *Tnfa* and mouse *H2-Aa*, *H2-Ab1*, *H2-Eb1*, and *H60a* (Qiagen QuantiTect Primer Assay). mRNA levels were normalized to *Gapdh* ($\Delta C_t = C_{t\text{gene of interest}} - C_{t\text{Gapdh}}$) and reported as relative mRNA expression ($\Delta\Delta C_t = 2^{-(\Delta C_{t\text{sample}} - \Delta C_{t\text{control}})}$) or fold change.

siRNA-mediated knockdown and gene transfection. Freshly isolated bone-marrow-derived CD11b⁺ myeloid cells or differentiated macrophages were transfected by electroporation using an AMAXA mouse macrophage nucleofection kit with 100 nM of siRNA or 2 µg *Pik3cg*CAAX or pcDNA control plasmid. Non-silencing (Ctrl_AllStars_1) siRNA and *Cebpb* (MmCebpb_4 and MmCebpb_6), and *Mtor* (Mm_Frap1_1 and Mm_Frap1_2) siRNAs were purchased from Qiagen. After transfection, cells were cultured for 36–48 h in RPMI containing 10% serum and 10 ng ml^{−1} mCSF (PeproTech) or polarized as described above.

ELISA assays. Whole tumours, CD11b⁺Gr1[−] cells, CD90.2⁺ cells, CD4⁺ cells and CD8⁺ cells isolated from LLC tumours were lysed in RIPA buffer and total protein concentrations were determined using a BCA protein assay (Pierce). Macrophage supernatants (100 µl) or 500 µg of total protein lysate from tumours were used in ELISAs to detect CCL2, TGFβ, IL1β, TNFα, IL6, IFNγ, IL10, IL12 and granzyme B

(ready set go ELISA, eBioscience). Protein expression was normalized to total volume (supernatants) or mg total protein (tumour lysates).

Quantitative colourimetric arginase determination. The QuantiChrom arginase assay kit (DARG-200, BioAssay Systems) was used to measure arginase activity in primary mouse bone-marrow-derived macrophages from wild-type and *Pik3cg*^{-/-} mice according to the manufacturer's instructions. For all conditions, cells were harvested and lysed in 10 mM Tris (pH 7.4) containing 1 μ M pepstatin A, 1 μ M leupeptin, and 0.4% (w/v) Triton X-100. Samples were centrifuged at 20,000g at 4°C for 10 min.

Transcription factor assays. To measure NF κ B and C/EBP β activation, TransAM NF κ B family and C/EBP transcription factor assay kits (43296 and 44196, Active Motif) were used according to the manufacturer's protocol. Briefly, wild-type and *Pik3cg*^{-/-} bone-marrow-derived macrophages were stimulated with LPS (100 ng ml⁻¹) or IL4 (20 ng ml⁻¹) and nuclear extracts were prepared in lysis buffer AM2 (Active Motif). Nuclear extracts were incubated with the immobilized consensus sequences and RelA, cRel or C/EBP β were detected using specific primary antibodies. Quantification was performed via colourimetric readout of absorbance at 450 nm.

Immunoblotting. IL4 and LPS macrophage cultures were solubilized in RIPA buffer containing protease and phosphatase inhibitors. Thirty micrograms of protein was electrophoresed on Biorad precast gradient gels and electroblotted onto PVDF membranes. Proteins were detected by incubation with 1:1,000 dilutions of primary antibodies, washed and incubated with goat anti-rabbit-HRP antibodies and detected after incubation with a chemiluminescent substrate. Primary antibodies directed against Akt (11E7), p-Akt (244F9), I κ B α (L35A5), I κ K β (D30C6), p-I κ K α / β (16A6), RelA (D14E12), pRelA (93H1), C/EBP β (#3087), p-CEBP β (#3082), IRAK1 (D51G7), TBK1 (D1B4) and PI3K γ (#4252) were from Cell Signaling Technology and pTBK1 (EPR2867(2)) was from Abcam.

In vitro cytotoxicity assay. CD90.2⁺ tumour-derived T cells were purified from LLC tumour-bearing wild-type and *Pik3cg*^{-/-} or TG100-115 and control treated mice and then co-incubated with LLC tumour cells (target cells) at 2.5:1, 5:1 and 10:1 ratios of T cells to tumour cells (2×10^3 LLC tumour cells per well) for 6 h. Target cell killing was assayed by collecting the supernatants from each well for measurement of the lactate dehydrogenase release (Cytotox96 non-radioactive cytotoxicity assay kit, Promega).

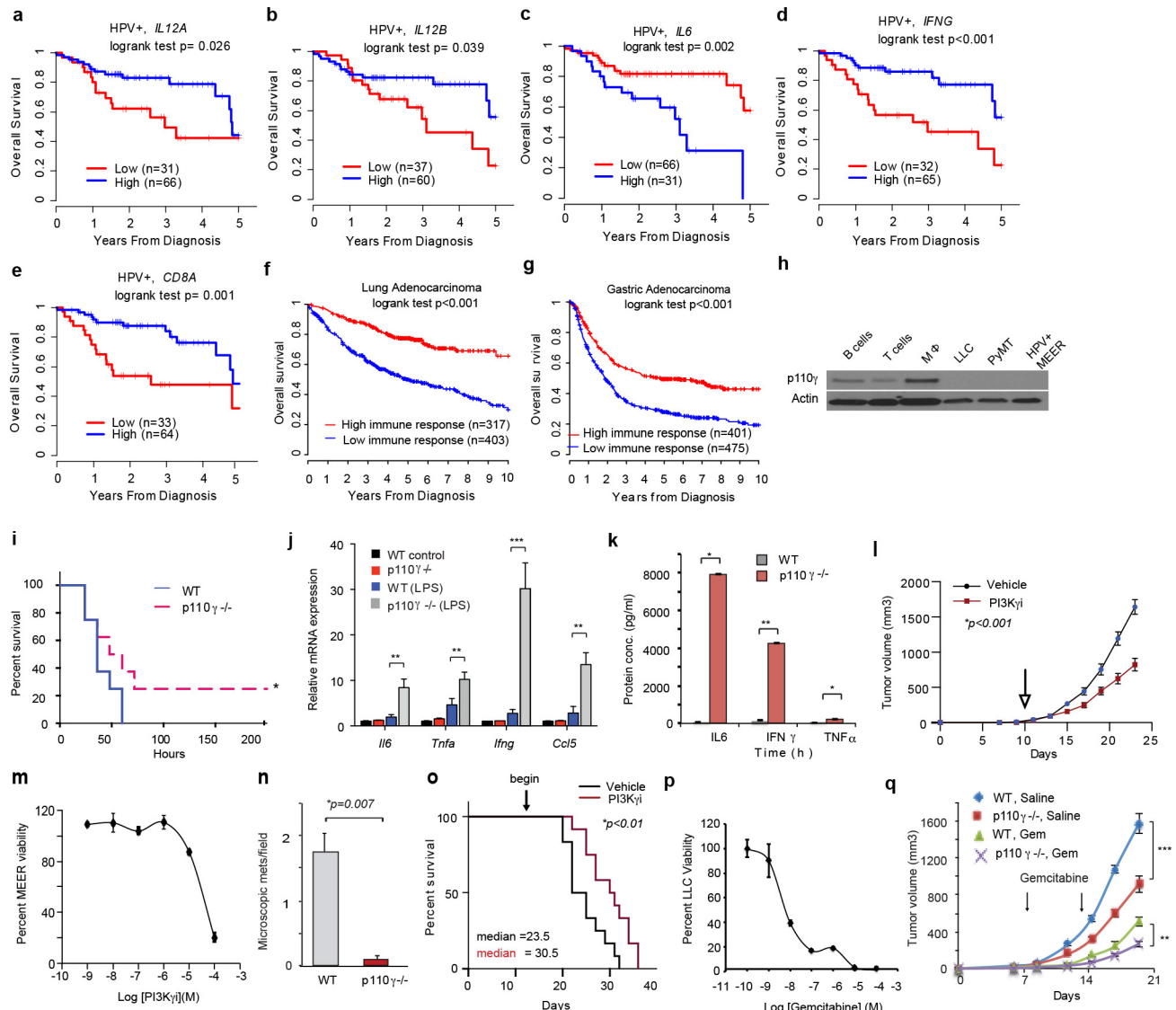
Immunohistochemistry. Tumour samples were collected and cryopreserved in OCT, sections (5 μ m) were fixed in 100% cold acetone, blocked with 8% normal goat serum for 2 h, and incubated anti-CD8 (53-6.7, 1:50 BD Biosciences) for 2 h at room temperature. Sections were washed three times with PBS and incubated with Alexa594-conjugated secondary antibodies. Slides were counterstained with

4',6-diamidino-2-phenylindole (DAPI) to identify nuclei. Immunofluorescence images were collected on a Nikon microscope (Eclipse TE2000-U) and analysed using Metamorph image capture and analysis software (Version 6.3r5, Molecular Devices). The detection of apoptotic cells was performed using a TUNEL-assay (ApopTag fluorescein *in situ* apoptosis detection kit, Promega) according to the manufacturer's instructions. Slides were washed and mounted in DAKO fluorescent mounting medium. Immunofluorescence images were collected on a Nikon microscope (Eclipse TE2000-U) and analysed with MetaMorph software (version 6.3r5) or SPOT software (version 4.6). Pixels per field or cell number per field were quantified in five 100 \times fields from ten biological replicates.

Statistics. Primary tumour samples with mRNA expression data were scored as above or below the median expression level, and tested for association with patient survival using a log-rank test at 5% significance. For studies evaluating the effect of drugs on tumour size, tumour dimensions were measured directly before the start of treatment, tumour volumes were computed and mice were randomly assigned to groups so that the mean volume \pm s.e.m. of each group was identical. A sample size of ten mice per group provided 80% power to detect a mean difference of 2.25 standard deviation (s.d.) between two groups (based on a two-sample *t*-test with two-sided 5% significance level). Sample sizes of 15 mice per group provided 80% power to detect one s.d. difference between two groups. Data were normalized to the standard (control). Analysis for significance was performed by one-way ANOVA with a Tukey's post-hoc test for multiple pairwise testing with more than two groups and by parametric or nonparametric Student's *t*-test when only two groups were compared. We used a two-sample *t*-test (two groups) and ANOVA (multiple groups) when data were normally distributed and a Wilcoxon rank sum test (two groups) when data were not normally distributed. All mouse studies were randomized and blinded; assignment of mice to treatment groups, tumour measurement and tumour analysis was performed by coding mice with randomly assigned mouse number, with the key unknown to operators until experiments were completed. In tumour studies for which tumour size was the outcome, mice removed from the study owing to health concerns were not included in endpoint analyses. All experiments were performed at least twice; *n* refers to biological replicates.

Data availability. RNA sequencing data can be accessed using numbers GSE58318 (*in vitro* macrophage samples) and GSE84318 (*in vivo* tumour and tumour-associated macrophages samples) at www.ncbi.nlm.nih.gov/geo.

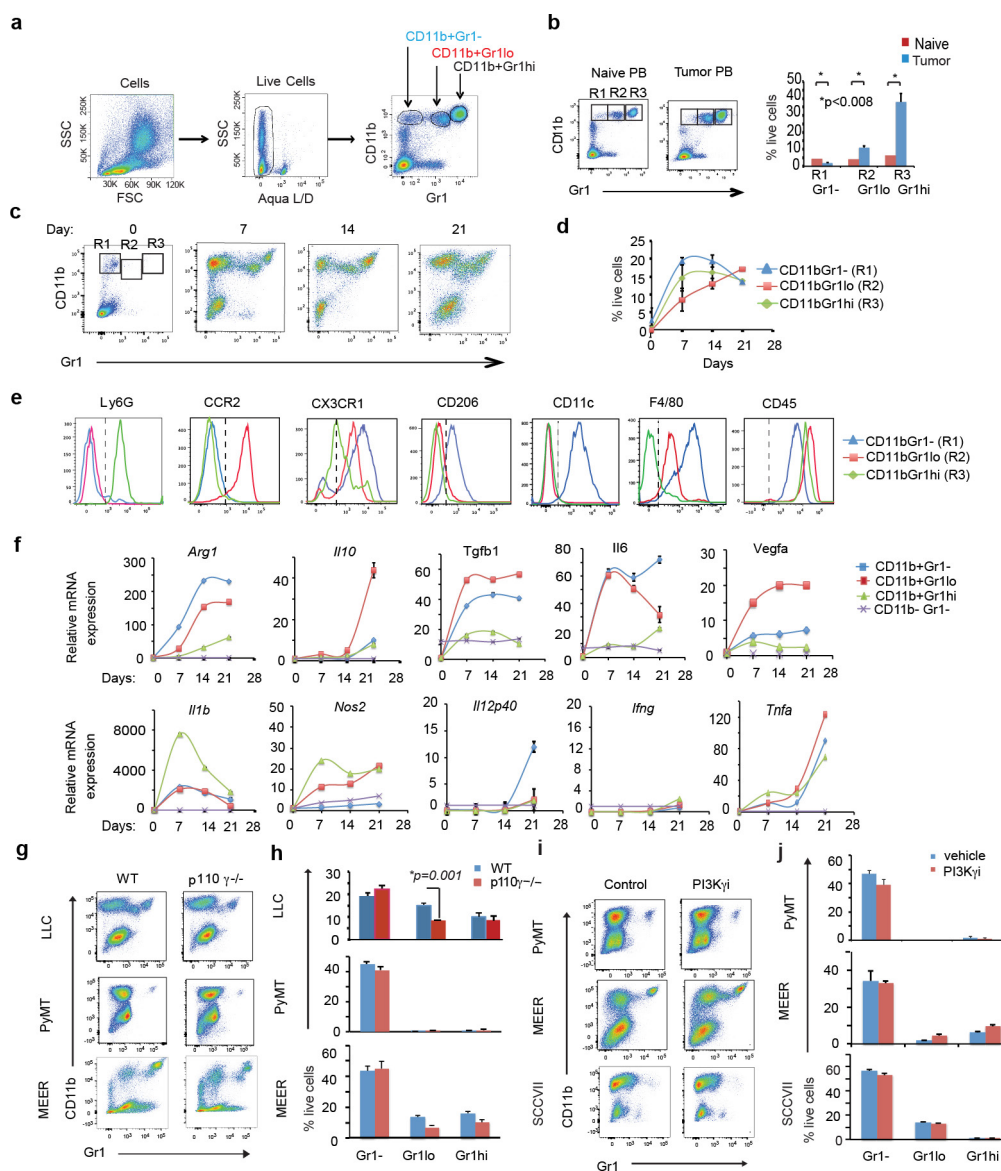
29. Györfy, B., Surowiak, P., Budczies, J. & Lánczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* **8**, e82241 (2013).



Extended Data Figure 1 | Pro-inflammatory gene expression signatures predict survival in cancer patients.

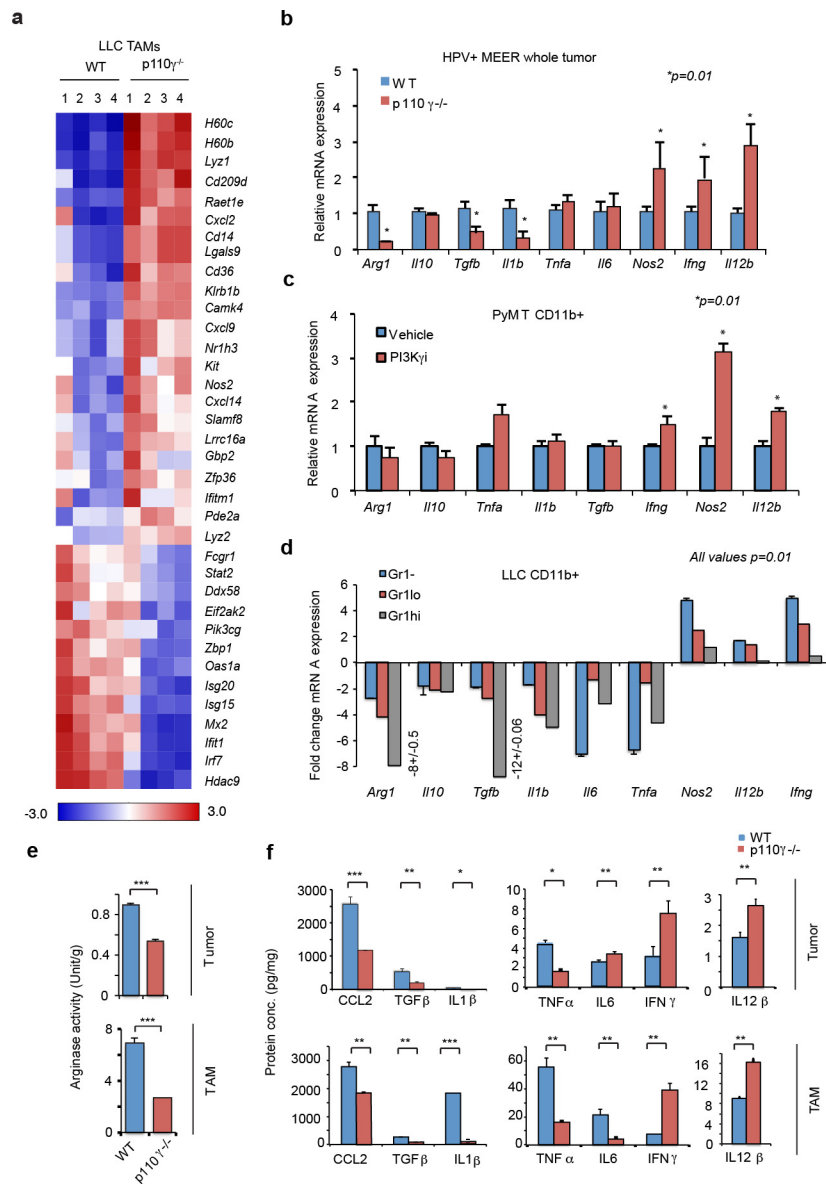
a–e, Association of *IL12A* ($P = 0.026$), *IL12B* ($P = 0.039$), *IFNG* ($P = 0.002$), *CD8A* ($P = 0.001$) and *IL6* ($P = 0.001$) with survival in 97 HPV⁺ HNSCC patients (log-rank test). **f**, Multivariate immune signature for 720 lung adenocarcinoma samples from patients from KM plotter cohorts ($P = 0.001$; log-rank test). **g**, Multivariate immune signature in 876 gastric cancer samples from patients from KM plotter cohorts ($P = 0.001$; log-rank test). **h**, Western blotting of PI3K γ in B cells, T cells, macrophages (M Φ) and LLC, PyMT and MEER tumour cells. **i**, Kaplan-Meier survival plot of wild-type (WT) and *Pik3cg^{-/-}* mice inoculated with LPS (endotoxin). $P = 0.05$, log-rank test. **j**, Pro-inflammatory cytokine mRNA expression in bone marrow from wild-type and *Pik3cg^{-/-}* LPS-injected mice. $n = 4$ biological replicates; ** $P < 0.001$; *** $P < 0.0001$; one-sided ANOVA with Tukey's post-hoc test. **k**, Circulating inflammatory cytokine levels in *Pik3cg^{-/-}*

and wild-type mice 24 h after endotoxin administration. $n = 4$ biological replicates; * $P < 0.01$; ** $P < 0.001$; one-sided ANOVA with Tukey's post-hoc test. **l**, Tumour volume of implanted HPV⁻ (SCCVII) carcinomas ($n = 15$ biological replicates) from vehicle or PI3K γ -inhibitor-treated mice. Arrow, start of drug treatment; $P = 0.001$; t -test. **m**, Dose-response of the effect of PI3K γ inhibitor IPI549 on *in vitro* MEER cell viability. **n**, Spontaneous PyMT lung metastases per high-power field (200 \times) in wild-type and *Pik3cg^{-/-}* mice. $n = 8$ biological replicates; $P = 0.007$; t -test. **o**, Kaplan-Meier survival plot of mice bearing orthotopic PyMT tumours treated with vehicle or the PI3K γ inhibitor IPI549 initiated as indicated by the arrow ($n = 10$). $P < 0.01$. **p**, *In vitro* LLC tumour cell survival in the presence of gemcitabine. **q**, Volume of LLC tumours implanted in wild-type and *Pik3cg^{-/-}* mice treated with saline or gemcitabine. $n = 10$ biological replicates; ** $P < 0.001$; *** $P < 0.0001$. All experiments were performed two or more times. **j–l**, **m**, **q**, Data are shown as mean \pm s.e.m.



Extended Data Figure 2 | Effect of PI3K γ inhibition on tumour inflammation. **a**, Gating strategy for flow cytometry analysis of myeloid cell populations in peripheral blood leukocytes. **b**, Representative flow cytometry analysis and quantification of myeloid cell populations in peripheral blood (PB) of naive and LLC tumour-bearing mice. $n=3$ biological replicates; $P < 0.008$; t -test. **c**, Flow cytometry analysis of myeloid cell populations on days 0, 7, 14 and 21 after subcutaneous inoculation with Lewis lung carcinoma cells ($n=3$ biological replicates). **d**, Quantification of populations from **c**. **e**, Flow cytometry analysis of Ly6G, CCR2, CX3CR1, CD206, CD11c, F4/80 and CD45 expression in myeloid cell populations from **c** ($n=3$ biological replicates). **f**, Relative immune response transcript levels \pm s.e.m. in tumour-derived myeloid

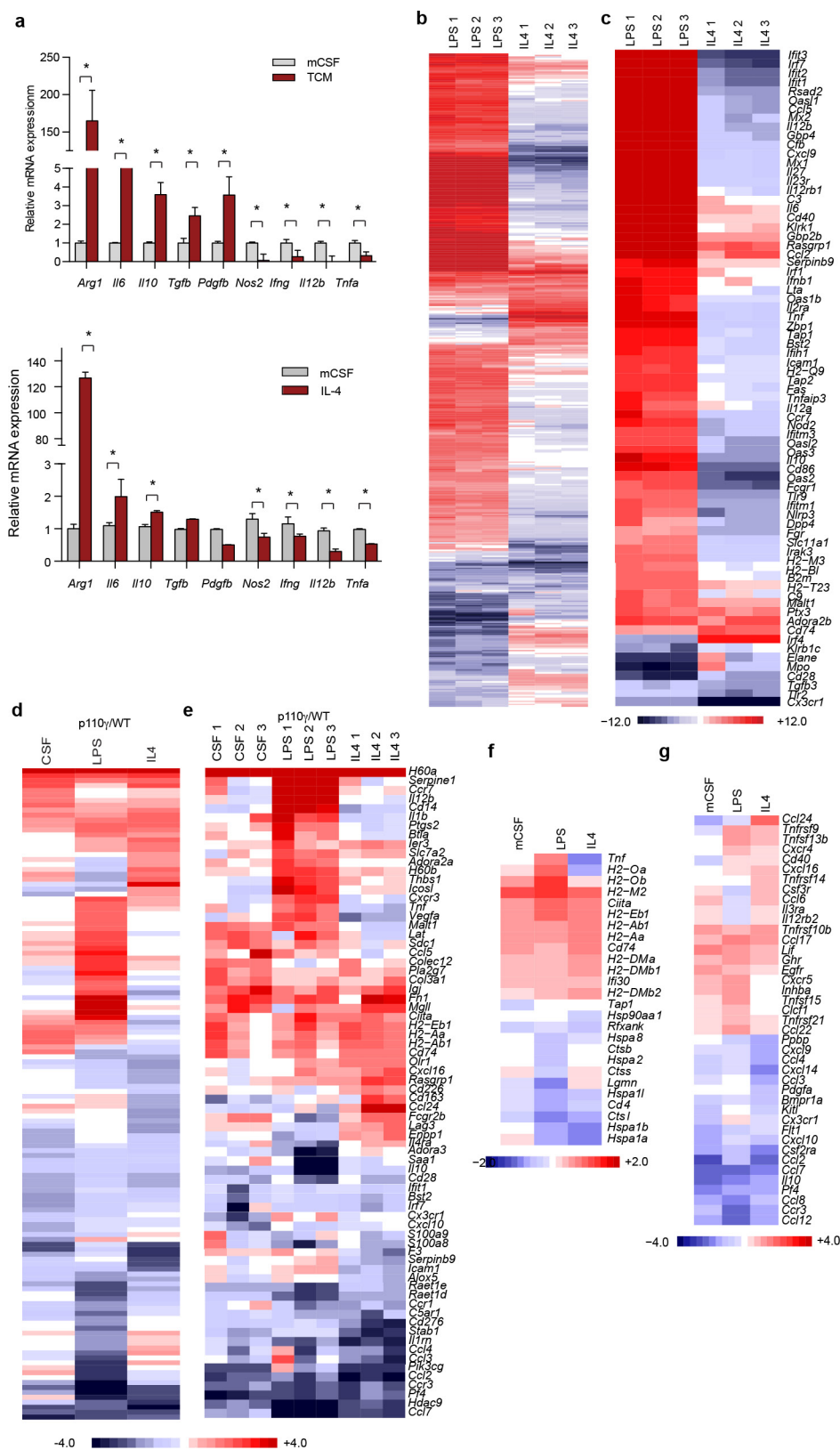
cells and tumour cells (CD11b⁺Gr1⁺ cells) isolated at day (d)0 ($n=3$), d7 ($n=5$), d14 ($n=3$) or d21 ($n=4$) after LLC cell inoculation ($P < 0.002$, d21 versus d0). n , biological replicates. **g** Flow cytometry analysis of CD11b⁺ myeloid cell populations in wild-type and *Pik3cg*^{-/-} LLC, PyMT and MEER tumours ($n=3$ biological replicates). **h**, Quantification of CD11b⁺ myeloid cell populations ($P=0.001$; t -test) from **g**. **i**, Flow cytometry analysis of CD11b⁺ myeloid cell populations in vehicle and PI3K γ -inhibitor-treated PyMT, MEER and SCCVII tumours ($n=3$ biological replicates). **j**, Quantification of CD11b⁺ myeloid cell populations from **i**. All experiments were performed two or more times. **b**, **d**, **h**, **j**, Data are shown as mean \pm s.e.m.



Extended Data Figure 3 | Effect of PI3K γ inhibition on TAM expression profile.

a, Heat map of differentially expressed immune response genes in TAMs isolated from LLC tumours from wild-type and *Pik3cg*^{-/-} mice (*n* = 3 biological replicates; local false discovery rate < 0.1) obtained by RNA sequencing. **b**, Relative mRNA expression of immune response factors in HPV⁺ HNSCC MEER tumours from *Pik3cg*^{-/-} and wild-type mice (*n* = 4 biological replicates), **P* = 0.01; *t*-test. **c**, Relative mRNA expression of immune response factors in CD11b⁺ myeloid cells isolated from PyMT tumours grown in vehicle or PI3K γ -inhibitor-treated mice (*n* = 4 biological replicates), **P* = 0.01; *t*-test. **d**, Fold change in mRNA expression in CD11b⁺Gr1⁻ (macrophage), CD11b⁺Gr1^{lo} (monocyte)

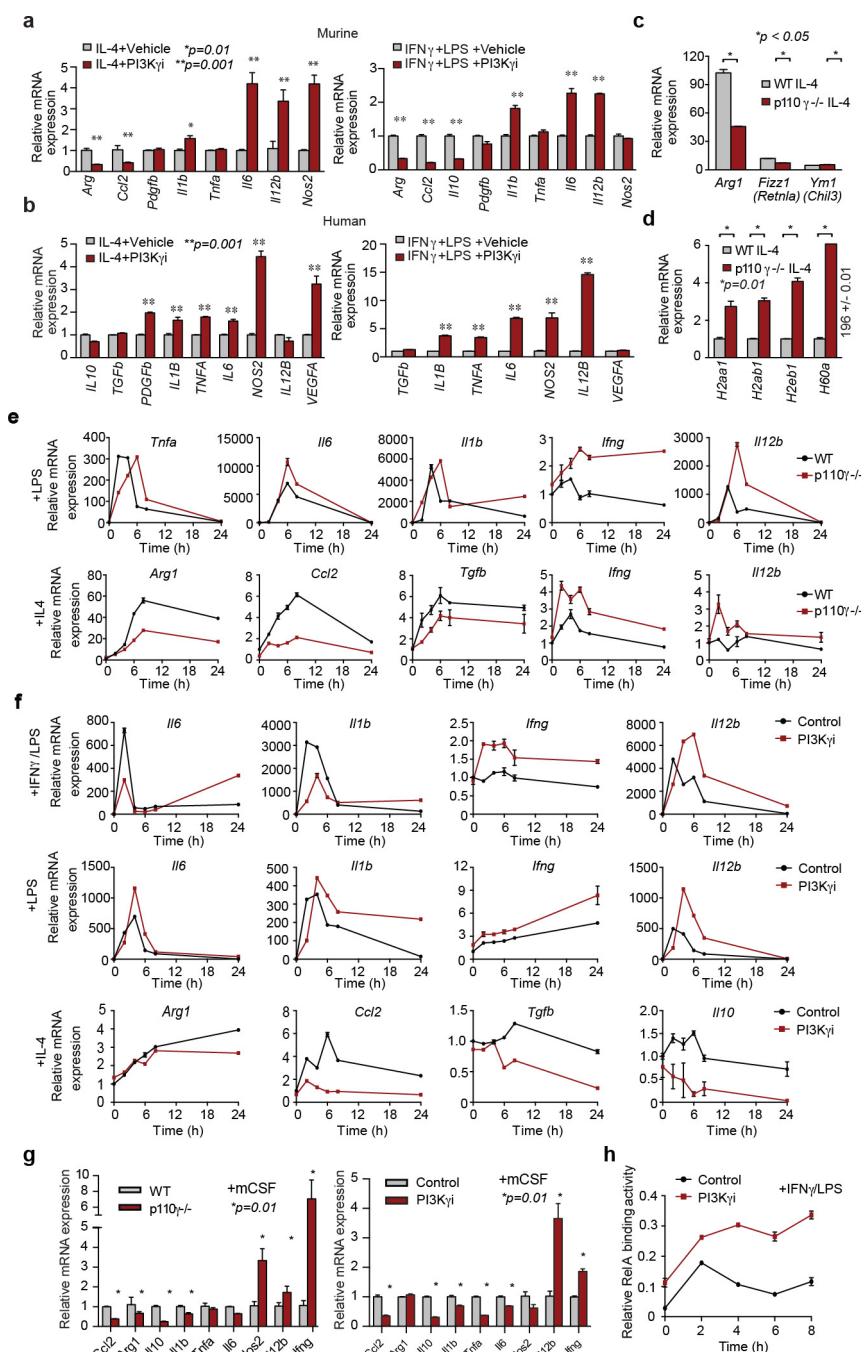
and CD11b⁺Gr1^{hi} (granulocyte) myeloid cells isolated from LLC tumours grown in *Pik3cg*^{-/-} mice (*n* = 5 biological replicates) and normalized to wild-type control. *n* = 5 biological replicates; *P* = 0.001; one-sided ANOVA with Tukey's post-hoc test. **e**, Arginase activity in tumours and TAMs isolated from LLC tumours grown in wild-type and *Pik3cg*^{-/-} mice. *n* = 4 biological replicates; ****P* < 0.0003; *t*-test. **f**, Protein concentration of cytokines in LLC tumours and TAMs from wild-type and *Pik3cg*^{-/-} mice. *n* = 4 biological replicates; **P* < 0.01; ***P* < 0.001; ****P* < 0.0001; *t*-test. All experiments were performed two or more times. **b–f**, Data are shown as mean \pm s.e.m.



Extended Data Figure 4 | See next page for caption.

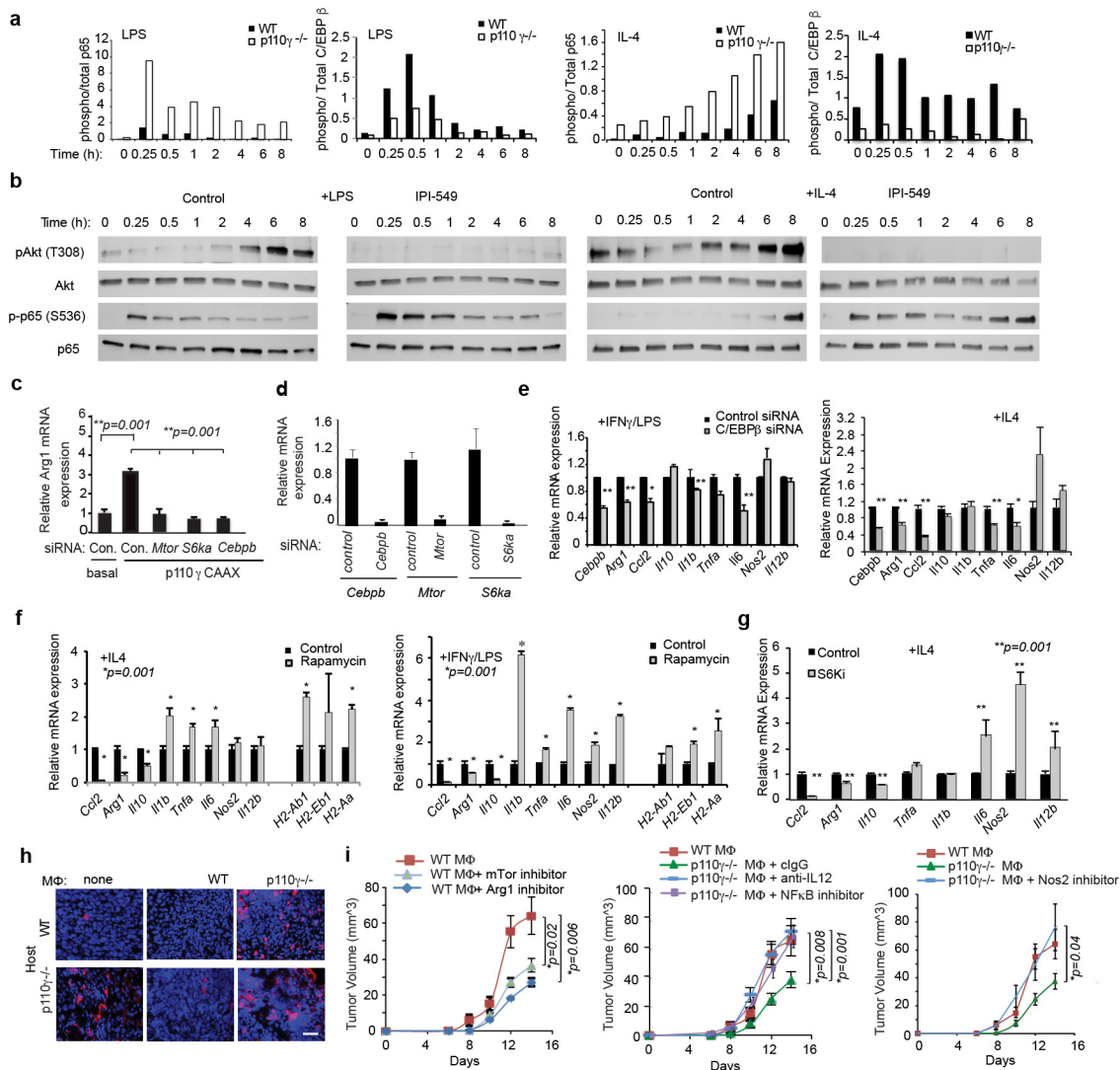
Extended Data Figure 4 | Effect of PI3K γ deletion on *in vitro* macrophage mRNA expression. **a**, Relative immune response mRNA expression in *Pik3cg*^{-/-} and wild-type (WT) mouse macrophages stimulated by IL4- or LLC-tumour-cell-conditioned medium as determined by RT-PCR. Data are shown as mean \pm s.e.m.; $n = 3$ biological replicates; $*P = 0.01$; t -test. **b**, Heat map of differentially expressed immune response transcripts in IL4- and IFN γ /LPS-polarized mouse macrophages obtained by RNA sequencing. $n = 3$ biological replicates; $P = 0.00001$. **c**, Heat map of select differentially expressed immune response transcripts in *in vitro* polarized mouse macrophages. $n = 3$ biological replicates; $P = 0.00001$. **d**, Heat map of immune response transcripts in mCSF-, IL4- and IFN γ /LPS-stimulated *Pik3cg*^{-/-} mouse

macrophages obtained by RNA sequencing and normalized to wild-type macrophages. $n = 3$ biological replicates; $P = 0.00001$. **e**, Heat map of select differentially expressed immune response transcripts in polarized *Pik3cg*^{-/-} mouse macrophages normalized to wild-type. $n = 3$ biological replicates; $P = 0.00001$. **f**, Heat map of differentially expressed antigen presentation and processing mRNAs in mCSF, IL4 and IFN γ /LPS-polarized *Pik3cg*^{-/-} mouse macrophages. $n = 3$ biological replicates; $P = 0.00001$. **g**, Heat map of differentially expressed chemokine and chemokine receptor mRNAs in polarized *Pik3cg*^{-/-} mouse macrophages. $n = 3$ biological replicates; $P = 0.00001$. All experiments were performed two or more times.



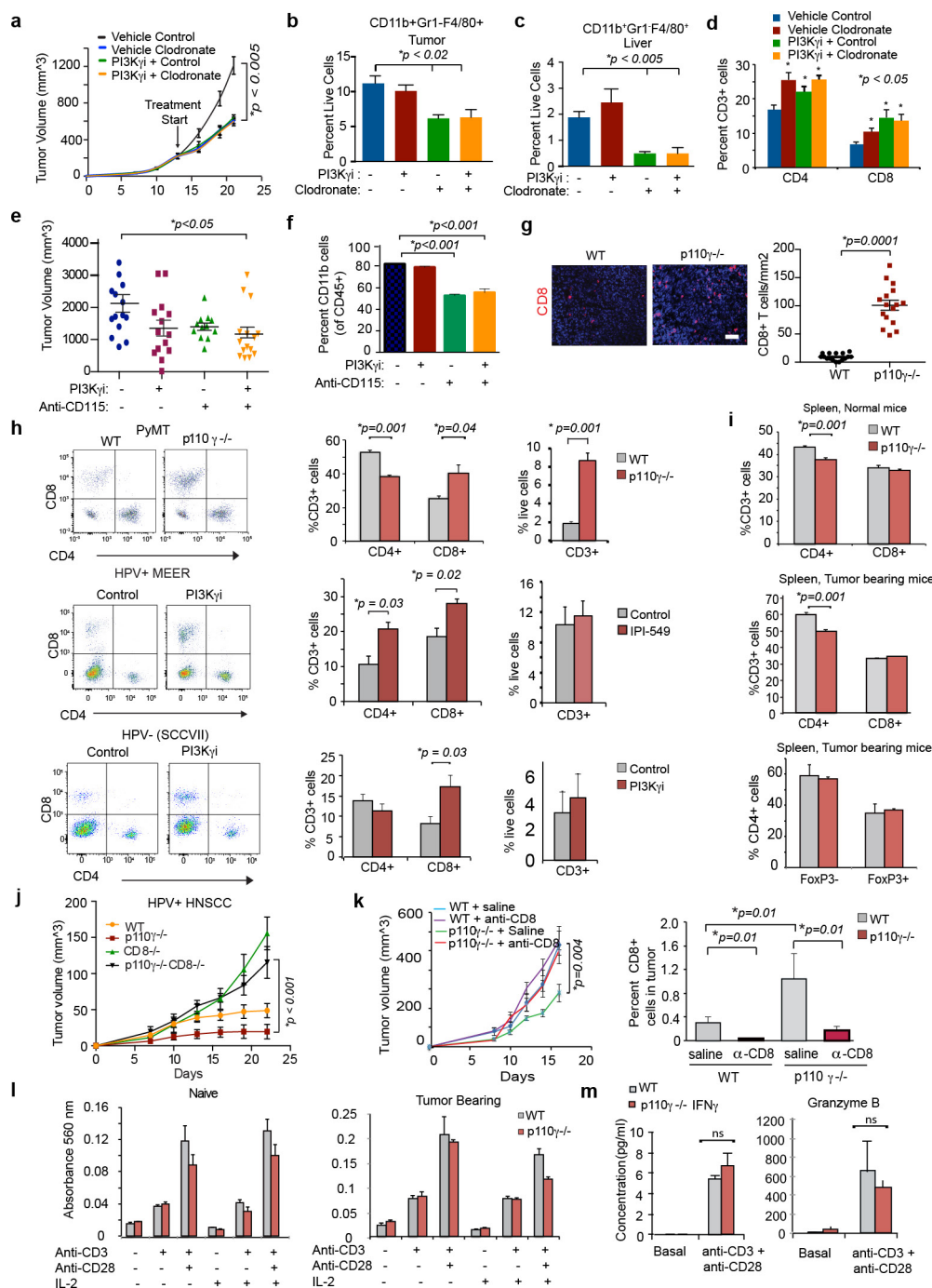
Extended Data Figure 5 | Effect of PI3K γ inhibition on mouse and human macrophage polarization. **a, b**, Relative mRNA expression of immune response transcripts in IL4 and IFN γ /LPS-stimulated vehicle and PI3K γ inhibitor (IPI-549)-treated mouse (**a**) and human (**b**) macrophages. $n = 3$ biological replicates; $*P = 0.01$; $**P = 0.001$; t -test. **c**, Relative mRNA expression of M2 macrophage markers (*Arg1*, *Retnla* (also known as *Fizz1*) and *Chil3*) in wild-type and *Pik3cg*^{-/-} IL4-stimulated macrophages ($n = 3$ biological replicates; $P = 0.05$; t -test). **d**, Relative RNA expression of MHC family members in wild-type and *Pik3cg*^{-/-} IL4-stimulated macrophages. $n = 3$ biological replicates; $P = 0.01$; t -test.

e, f, mRNA expression of cytokines over time in IFN γ /LPS, LPS and IL4 stimulated wild-type or *Pik3cg*^{-/-} (**e**) and vehicle- or PI3K γ -inhibitor- (IPI-549)-treated (**f**) macrophages ($n = 3$ biological replicates). **g**, Relative mRNA expression in mCSF-stimulated wild-type or *Pik3cg*^{-/-} and IPI-549- or vehicle-treated macrophages. $n = 3$ biological replicates; $P = 0.01$; t -test. **h**, Relative nuclear RelA DNA binding activity in IFN γ /LPS stimulated wild-type and *Pik3cg*^{-/-} macrophages. $n = 3$ biological replicates; $P = 0.01$; t -test. All experiments were performed two or more times.



Extended Data Figure 6 | Mechanism of PI3K γ -mediated gene expression regulation. **a**, Relative levels of phospho/total p65 and phospho/total C/EBP β in LPS- and IL4-stimulated wild-type and *Pik3cg*^{-/-} macrophages. **b**, Immunoblotting to detect pThr308Akt, total Akt, phospho-p65 and total p65 in LPS- and IL4- stimulated macrophages that were treated with vehicle or the PI3K γ inhibitor IPI-549. **c**, Relative *Arg1* mRNA expression in myeloid cells transfected with constitutively active, membrane-targeted PI3K γ (*Pik3cg*CAAX) and *Mtor*, *S6ka*, *Cebpb* or control siRNA. *n* = 3 biological replicates; *P* = 0.001, *t*-test. **d**, Validation of mRNA expression in macrophages expressing siRNAs from **c** (*n* = 3 biological replicates). **e**, Effect on cytokine mRNA expression in wild-type macrophages transfected with *Cebpb*, *Mtor* or *S6ka* siRNA.

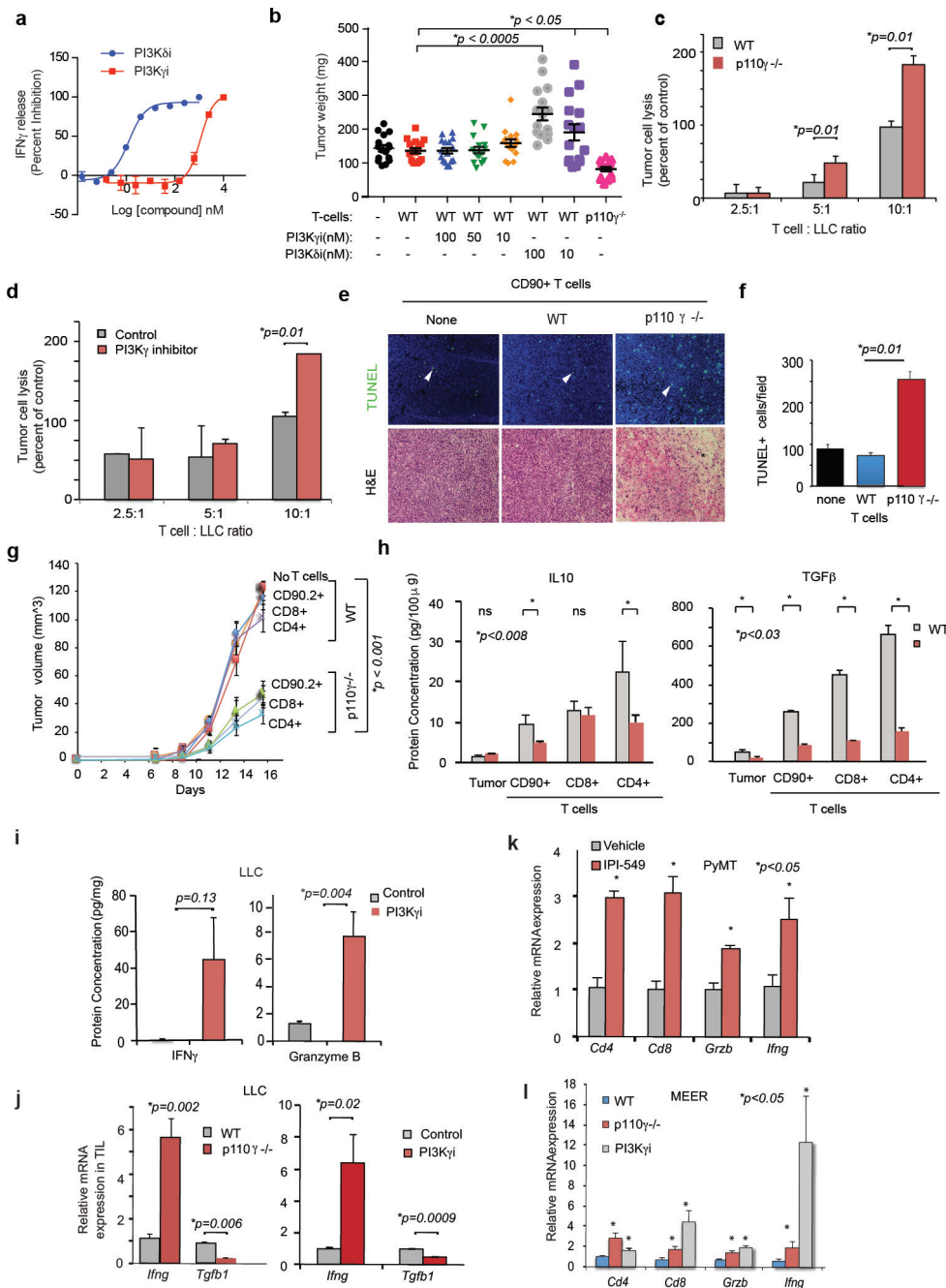
n = 3 biological replicates; **P* = 0.01; ***P* = 0.001; *t*-test. **f**, **g**, Cytokine mRNA expression in macrophages treated with rapamycin (**f**) or the S6K inhibitor PF4708671 (**g**) (*n* = 3 biological replicates, *P* = 0.001, *t*-test). **h**, Immunofluorescence images of CD8⁺ T cells in 10 μm tumour sections from animals in Fig. 3b. **c**, **i**, Tumour volumes from tumour cells mixed with wild-type TAMs pretreated with the mTOR inhibitor rapamycin or the arginase inhibitor nor-NOHA and *Pik3cg*^{-/-} TAMs pretreated with anti-IL12 or isotype-matched control antibody (cIgG), the IκKβ inhibitor MLB120 or the NOS2 inhibitor 1400W dihydrochloride (*n* = 10 biological replicates; *P* ≤ 0.04, one-sided ANOVA with Tukey's post-hoc test). All experiments were performed two or more times. **c**–**g**, **i**, Data are shown as mean ± s.e.m.



Extended Data Figure 7 | No direct effect of PI3K γ inhibition on T cells.

a, Volumes of LLC tumours treated with vehicle + control liposomes, PI3K γ inhibitor (IPI-549) + control liposomes, clodronate liposomes + vehicle and PI3K γ inhibitor + clodronate liposomes. $n = 10$ biological replicates; $P = 0.005$; one-sided ANOVA with Tukey's post-hoc test. **b**, Quantification of F4/80⁺ macrophages in tumours from **a**. $n = 3$ biological replicates; $P = 0.02$; one-sided ANOVA with Tukey's post-hoc test. **c**, Quantification of F4/80⁺ macrophages in livers from **a**. $n = 3$ biological replicates; $P < 0.005$; one-sided ANOVA with Tukey's post-hoc test. **d**, Quantification of T cells in tumours from **a**. $n = 3$ biological replicates; $P < 0.05$; one-sided ANOVA with Tukey's post-hoc test. **e**, Volumes of CT26 tumours treated with vehicle + cIgG, PI3K γ inhibitor (IPI-549) + cIgG, anti-CD115 + vehicle and PI3K γ inhibitor + anti-CD115. $n = 15$ biological replicates; $P = 0.05$; one-sided ANOVA with Tukey's post-hoc test. **f**, Quantification of CD11b⁺ myeloid cells in tumours from **e**. $n = 5$ biological replicates; $P < 0.001$; one-sided ANOVA with Tukey's post-hoc test. **g**, Images and quantification of CD8⁺ T cells in wild-type and *Pik3cg*^{-/-} LLC tumours by immunohistochemistry. $n = 5$

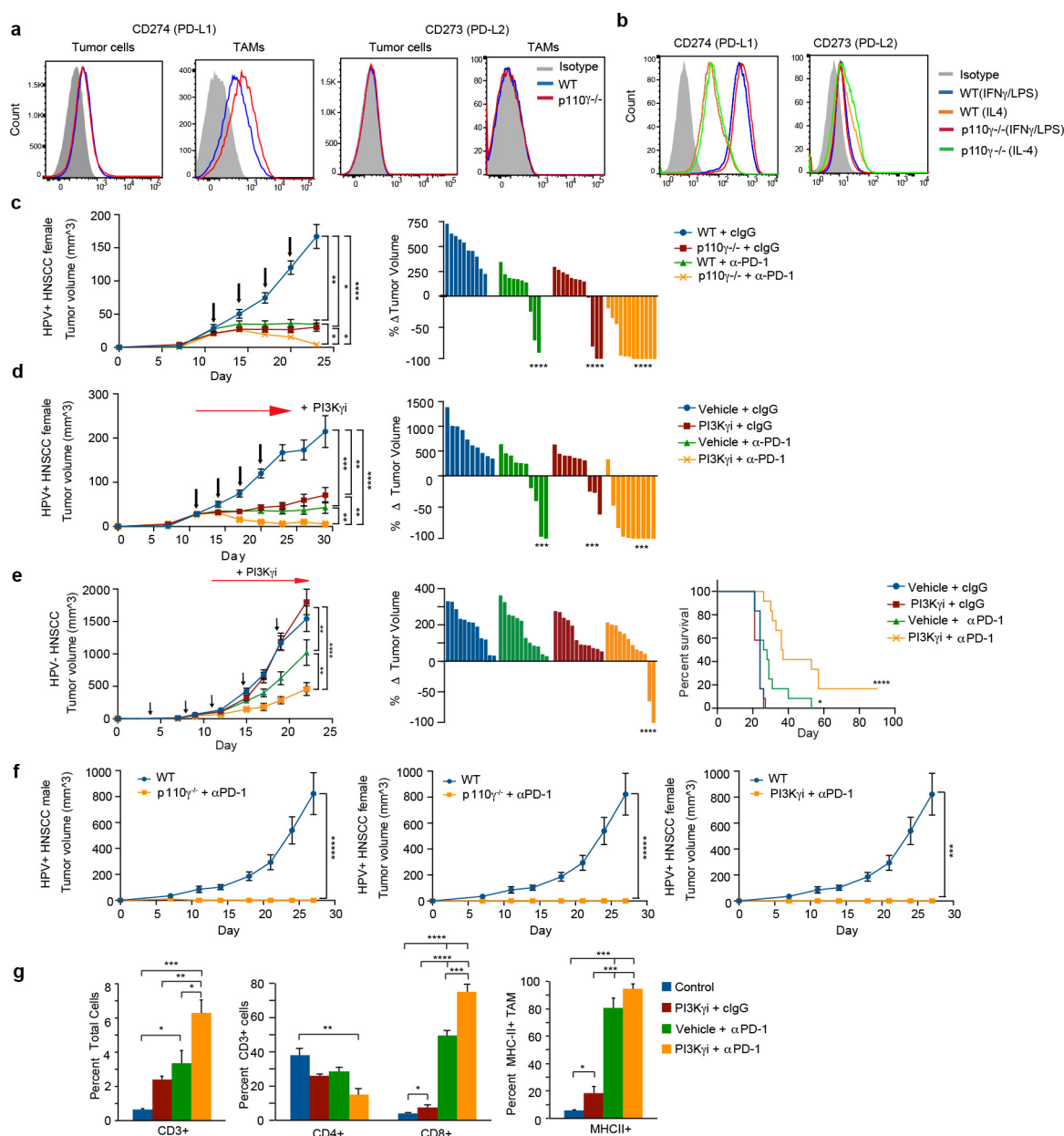
biological replicates; $P = 0.001$; one-sided ANOVA with Tukey's post-hoc test. **h**, Flow cytometry analysis and quantification of T cell populations in tumours from wild-type and *Pik3cg*^{-/-} or IPI-549-treated mice. ($n = 3$ biological replicates; $P < 0.05$; t -test). **i**, Quantification of T cells in spleens of naive and LLC tumour-bearing wild-type and *Pik3cg*^{-/-} mice. $n = 3$ biological replicates; $P = 0.001$; t -test. **j**, Volumes of LLC lung tumours from wild-type, *Pik3cg*^{-/-}, *CD8*^{-/-} and *CD8*^{-/-}, *Pik3cg*^{-/-} mice. $n = 12$ biological replicates; $P < 0.001$; one-sided ANOVA with Tukey's post-hoc test. **k**, LLC tumour volume from wild-type and *Pik3cg*^{-/-} mice treated with anti-CD8 antibodies or control ($n = 10$ biological replicates; $P = 0.004$; one-sided ANOVA with Tukey's post-hoc test) and per cent CD8⁺ T cells out of CD3⁺ T cells in these tumours ($n = 3$ biological replicates; $P = 0.01$; t -test). **l**, *In vitro* proliferation of T cells (mean \pm s.e.m. absorbance at 560 nm) isolated from naive and LLC tumour-bearing wild-type and *Pik3cg*^{-/-} mice ($n = 3$ biological replicates). **m**, IFN γ and granzyme B protein expression in T cells from **l** ($n = 3$ biological replicates). All data are shown as mean \pm s.e.m. and all experiments were performed two or more times.



Extended Data Figure 8 | PI3K γ inhibition relieves T cell exhaustion.

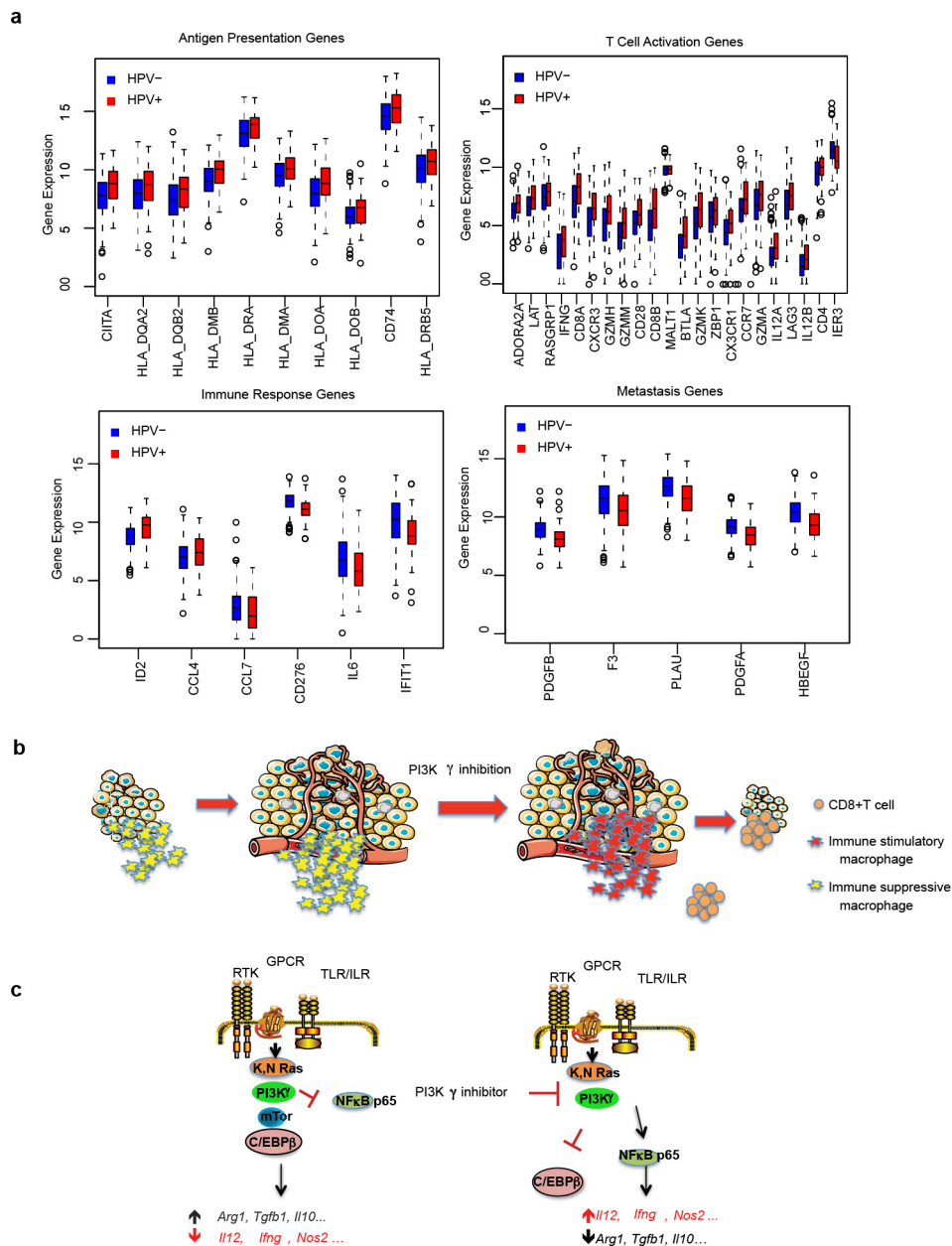
a, Expression of IFN γ in activated human T cells treated with PI3K γ and PI3K δ inhibitors. Data are shown as mean \pm s.d.; $n = 2$ biological replicates. **b**, Tumour weights derived from a mixture of LLC cells and wild-type or *Pik3cg*^{-/-} tumour-derived T cells or wild-type T cells pre-incubated with 10 or 100 nM PI3K γ (IPI-549) and PI3K δ (Cal101) inhibitors before implantation. $n = 16$ biological replicates; $P = 0.005$ (*Pik3cg*^{-/-}); $P = 0.05$ (PI3K γ); one-sided ANOVA with Tukey's post-hoc test. **c**, **d**, LLC tumour cell cytotoxicity induced by T cells isolated from LLC tumours from wild-type and *Pik3cg*^{-/-} (**c**) or control- and PI3K γ -inhibitor-treated (**d**) mice. $n = 3$ biological replicates; $*P < 0.001$; t -test. **e**, Images of TUNEL and haematoxylin and eosin stained tumours implanted with WT, *Pik3cg*^{-/-} or no T cells from tumours shown in Fig. 3h. **f**, Quantification of TUNEL⁺ cells in tumour sections from **e**. $n = 10$ biological replicates; $P = 0.01$; t -test. **g**, Tumour volumes in wild-type mice of tumours derived from LLC tumour cells mixed 1:1 with CD90.2⁺, CD4⁺ and CD8⁺ T cells or no T cells from wild-type or

Pik3cg^{-/-} tumour-bearing mice. $n = 8$ biological replicates; $P = 0.001$; one way ANOVA with Tukey's post-hoc test. **h**, mRNA expression of IL10 ($P = 0.008$; t -test) and TGF β ($P = 0.03$, t -test) protein expression in lysates from tumour and CD90.2⁺, CD8⁺ and CD4⁺ T cells isolated from LLC tumours grown in wild-type and *Pik3cg*^{-/-} mice ($n = 3$ biological replicates). **i**, IFN γ ($P = 0.13$, t -test) and granzyme B ($P = 0.004$, t -test) protein expression in PI3K γ -inhibitor- or control-treated LLC tumours ($n = 3$ biological replicates). **j**, *Ifng* and *Tgfb1* mRNA expression in T cells isolated from LLC tumours grown in wild-type and *Pik3cg*^{-/-} or control- and PI3K γ -inhibitor-treated mice. $n = 3$ biological replicates; $P = 0.05$, t -test). **k**, Relative mRNA expression of *Cd4*, *Cd8*, *Gzmb* and *Ifng* in control- and PI3K γ -inhibitor-treated PyMT tumours. $n = 3$ biological replicates; $P = 0.05$, t -test. **l**, Relative mRNA expression of *Cd4*, *Cd8*, *Gzmb* and *Ifng* in wild-type and *Pik3cg*^{-/-} and PI3K γ -inhibitor-treated HPV⁺ MEER tumours ($n = 3$ biological replicates, t -test). All experiments were performed two or more times. **b–d**, **f**, **g–j**, **l**, Data are shown as mean \pm s.e.m.



Extended Data Figure 9 | PI3K γ role in the macrophage-mediated tumour immune response. **a, b**, Flow cytometry analysis of PD-L1 and PD-L2 expression in tumour cells and TAMs from wild-type and *Pik3cg*^{-/-} LLC tumours (**a**) and wild-type and *Pik3cg*^{-/-} *in vitro* cultured IFN γ /LPS- and IL4-stimulated macrophages (**b**) ($n = 3$ biological replicates). **c**, HPV⁺ HNSCC tumour volume in female wild-type or *Pik3cg*^{-/-} mice that were treated with anti-PD-1 or isotype-matched antibody (cIgG), as indicated by arrows and per cent change in tumour volumes between days 11 and 23. $n = 10$ biological replicates; * $P = 0.01$; ** $P = 0.001$; *** $P = 0.0001$; **** $P = 0.00001$; one-sided ANOVA with Tukey's post-hoc test). **d**, HPV⁺ HNSCC tumour volume in female wild-type mice that were treated with PI3K γ inhibitor (2.5 mg kg⁻¹ TG100-115 twice per day) in combination with anti-PD-1 or isotype-matched antibody (cIgG), as indicated by arrows, and per cent change in tumour volumes between days 11 and 29. $n = 10$ biological replicates; * $P = 0.01$; ** $P = 0.001$; *** $P = 0.0001$; **** $P = 0.00001$, one-sided ANOVA with Tukey's post-hoc test). **e**, HPV⁺

HNSCC tumour volume in mice that were treated with PI3K γ inhibitor (2.5 mg kg⁻¹ TG100-115 twice per day) in combination with anti-PD-1 or cIgG, as indicated by arrows, per cent change in tumour volumes between days 19 and 26 and survival of treated mice. $n = 10$ biological replicates; * $P = 0.01$; ** $P = 0.001$; *** $P = 0.0001$; **** $P = 0.00001$; one-sided ANOVA with Tukey's post-hoc test). **f**, Tumour volume in HPV⁺ mice that had previously cleared HPV⁺ tumours and that were re-challenged with new HPV⁺ tumours ($n = 7$ –12 biological replicates) compared to wild-type mice newly implanted with HPV⁺ tumours ($n = 5$ biological replicates). *** $P = 0.0001$; **** $P = 0.00001$; one-sided ANOVA with Tukey's post-hoc test). **g**, Per cent CD3⁺, CD4⁺ and CD8⁺ T cells and MHCII⁺ macrophages from Fig. 4i. $n = 3$ biological replicates; * $P = 0.05$; ** $P = 0.005$; *** $P = 0.0005$; **** $P = 0.00005$; one-sided ANOVA with Tukey's post-hoc test. All experiments were performed two or more times. **c–g**, Data are shown as mean \pm s.e.m.



Extended Data Figure 10 | PI3K γ promotes immune suppression.

a, Comparison of median gene expression between HPV⁺ ($n = 97$) and HPV⁻ ($n = 423$) cohorts indicating HPV⁻ samples had significantly ($P < 0.05$, log-rank test) lower expression of adaptive immune genes and higher expression of immune suppressive and/or pro-metastasis genes. Blue, HPV⁻ samples; red, HPV⁺ samples. **b**, Model depicting the effect of PI3K γ inhibition on tumour immune suppression. PI3K γ inhibition converts tumour-associated macrophages into pro-inflammatory

macrophages that promote a CD8⁺ T cell response that suppresses tumour growth. **c**, Model depicting the PI3K γ signalling pathway in macrophages. PI3K γ activation attenuates NF κ B activation and promotes mTOR-dependent C/EBP β activation, leading to expression of immune suppressive factors and tumour growth. By contrast, PI3K γ inhibition inhibits C/EBP β and stimulates NF κ B, leading to altered expression of pro-inflammatory immune response cytokines.

Overcoming resistance to checkpoint blockade therapy by targeting PI3K γ in myeloid cells

Olivier De Henau¹, Matthew Rausch², David Winkler², Luis Felipe Campesato¹, Cailian Liu¹, Daniel Hirschhorn-Cymerman¹, Sadna Budhu¹, Arnab Ghosh¹, Melissa Pink², Jeremy Tchaicha², Mark Douglas², Thomas Tibbitts², Sujata Sharma², Jennifer Proctor², Nicole Kosmider², Kerry White², Howard Stern², John Soglia², Julian Adams², Vito J. Palombella², Karen McGovern², Jeffery L. Kutok², Jedd D. Wolchok^{1,3§} & Taha Merghoub^{1§}

Recent clinical trials using immunotherapy have demonstrated its potential to control cancer by disinhibiting the immune system. Immune checkpoint blocking (ICB) antibodies against cytotoxic-T-lymphocyte-associated protein 4 or programmed cell death protein 1/programmed death-ligand 1 have displayed durable clinical responses in various cancers¹. Although these new immunotherapies have had a notable effect on cancer treatment, multiple mechanisms of immune resistance exist in tumours. Among the key mechanisms, myeloid cells have a major role in limiting effective tumour immunity^{2–4}. Growing evidence suggests that high infiltration of immune-suppressive myeloid cells correlates with poor prognosis and ICB resistance^{5,6}. These observations suggest a need for a precision medicine approach in which the design of the immunotherapeutic combination is modified on the basis of the tumour immune landscape to overcome such resistance mechanisms. Here we employ a pre-clinical mouse model system and show that resistance to ICB is directly mediated by the suppressive activity of infiltrating myeloid cells in various tumours. Furthermore, selective pharmacologic targeting of the gamma isoform of phosphoinositide 3-kinase (PI3K γ), highly expressed in myeloid cells, restores sensitivity to ICB. We demonstrate that targeting PI3K γ with a selective inhibitor, currently being evaluated in a phase 1 clinical trial (NCT02637531), can reshape the tumour immune microenvironment and promote cytotoxic-T-cell-mediated tumour regression without targeting cancer cells directly. Our results introduce opportunities for new combination strategies using a selective small molecule PI3K γ inhibitor, such as IPI-549, to overcome resistance to ICB in patients with high levels of suppressive myeloid cell infiltration in tumours.

Tumour-associated myeloid cells (TAMCs) constitute a major component of the tumour microenvironment. Although some controversies persist in the precise description of the distinct subsets of this heterogeneous population, it is accepted that these cells promote tumour immunosuppression⁷. Recent studies support their contribution to the suppression of T cell function, which is not abolished by the use of ICB^{8–11}. To understand the association between resistance to ICB and myeloid cell infiltration, we compared multiple mouse tumour models treated with ICB. We show that mice bearing 4T1 breast carcinoma are resistant to anti-PD-1 (programmed cell death protein 1) or anti-CTLA4 (cytotoxic-T-lymphocyte-associated protein 4) therapy (Fig. 1a and Extended Data Fig. 1). We observe that myeloid cells (CD11b⁺) constitute the majority of CD45⁺ tumour-infiltrating leukocytes (TILs) in this model (Fig. 1b). This correlates with reduced CD8⁺ T cell infiltration and cytolytic function (Fig. 1b, c). By contrast, B16-F10 melanoma tumours, which are more responsive to ICB (Fig. 1a and Extended Data Fig. 1), exhibit less myeloid cell infiltration

but contain more activated CD8⁺ T cells (Fig. 1b, c). Additionally, CD8⁺ T cells express more granzyme B in the B16-F10 model. They also express higher levels of PD-1 and CTLA4 (Fig. 1c, data not shown), which might explain their sensitivity to ICB. Furthermore, myeloid cells from 4T1 tumours or spleens suppress proliferation of T cells to a greater extent compared to myeloid cells from B16-F10 models (Fig. 1d and Extended Data Fig. 1b). These data suggest that TAMCs have varying phenotypes and are more suppressive in ICB-resistant tumours. Tumour-derived soluble factors such as granulocyte-macrophage colony-stimulating factor (GM-CSF) help shape the tumour microenvironment by promoting myelopoiesis and recruitment of suppressive myeloid cells^{12,13}. To directly assess the ability of suppressive myeloid cells to induce resistance to ICB in the B16 melanoma, we used B16-F10 cells transduced with a GM-CSF expression construct (B16-GM-CSF)¹⁴. Tumours in this B16 model (B16-GM-CSF) become infiltrated by suppressive TAMCs and lose sensitivity to ICB compared to B16-F10 controls (Fig. 1a–d and Extended Data Fig. 1), indicating the critical role of suppressive myeloid cells in ICB resistance.

Kaneda *et al.*¹⁵ and other previous studies^{16,17} have shown that PI3K γ is highly expressed in myeloid cells and promotes migration and production of inflammatory mediators. Mice lacking the p110 γ subunit of PI3K γ show reduced tumour growth and immunosuppressive TAMCs¹⁸. We reasoned that pharmacological PI3K γ inhibition might lead to an effective anti-tumour immune response and subvert ICB resistance caused by suppressive myeloid cells. To test this hypothesis, we used IPI-549, a selective PI3K γ inhibitor in multiple tumour models¹⁹ (Extended Data Fig. 2a). IPI-549 treatment alone led to tumour growth inhibition in 4T1, B16-GMCSF, MC38, CT26, and LLC tumour models (Fig. 2a and Extended Data Fig. 2b). Lack of activity in tumours with few suppressive TAMCs (B16-F10, Fig. 2a) suggests PI3K γ inhibition affects myeloid cells and has no direct effect on tumour cells or other TILs. We also observe a significant reduction in lung metastasis after IPI-549-treatment (Fig. 2b). TIL quantification shows no consistent differences in the total myeloid (CD11b⁺) or macrophage (CD11b⁺F4/80⁺) cell populations in either IPI-549- or vehicle-treated 4T1 or B16-GMCSF tumours (Fig. 2c). The effects of IPI-549 on TAMC subsets were comparable to those seen in *Pik3cg*^{−/−} mice^{15,18}. PI3K γ inhibition switches the activation of macrophages from an immunosuppressive M2-like (CD11b⁺F4/80⁺CD206⁺) phenotype to a more inflammatory M1-like (CD11b⁺F4/80⁺MHCII⁺) state (Fig. 2c and Extended Data Fig. 2c). We further tested RNA expression of M1 and M2 markers in 4T1 and B16-GMCSF tumours after IPI-549. The expression of prototypic M2 markers (TGF β , Arg-1, IDO) are reduced, while M1 markers (IL-12, iNOS) are higher in IPI-549-treated tumours (Fig. 2d). Given that CD11b⁺F4/80⁺ macrophages

¹Memorial Sloan Kettering Cancer Center, Parker Institute for Cancer Immunotherapy and Swim Across America/Ludwig Collaborative Laboratory, New York, New York 10065, USA. ²Infinity Pharmaceuticals, Inc., Cambridge, Massachusetts 02139, USA. ³Weill Cornell Medical and Graduate Schools, New York, New York 10065, USA.

§These authors jointly supervised this work.

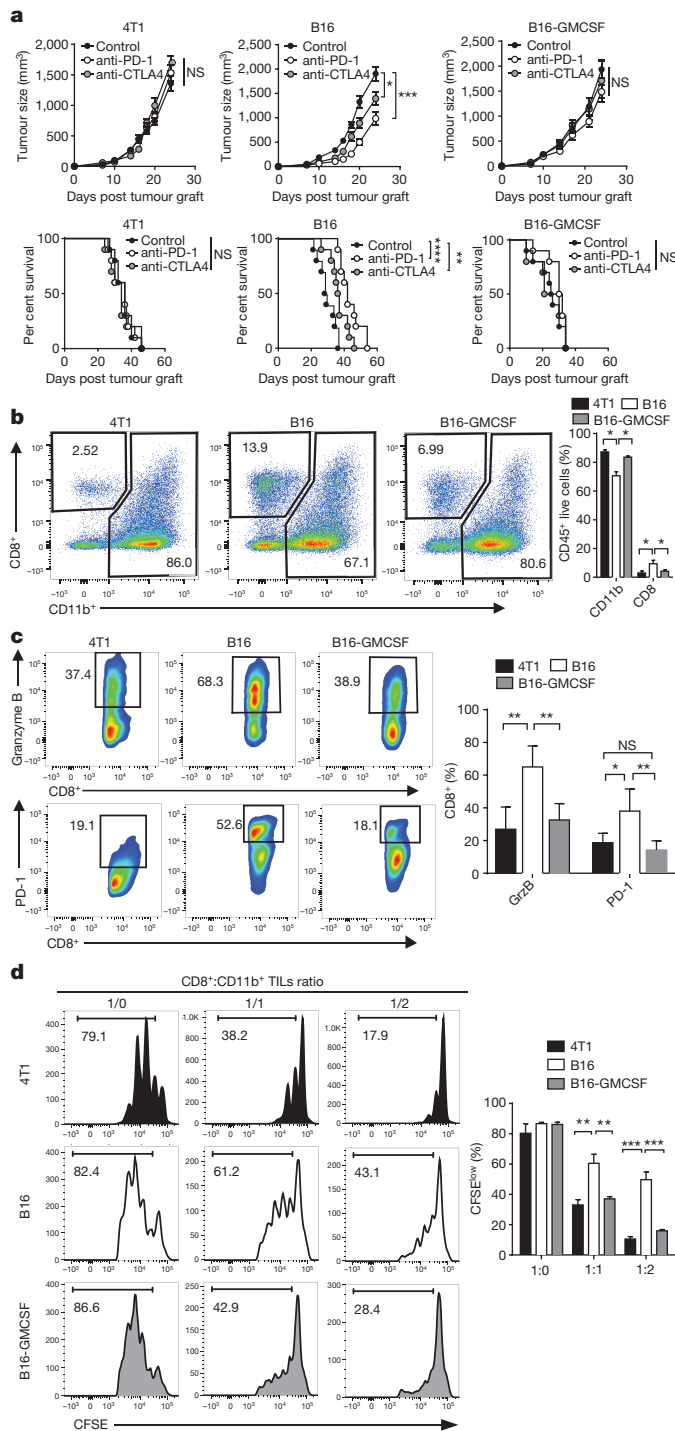


Figure 1 | Resistance to checkpoint blockade is associated with suppressive myeloid cells infiltration in tumour microenvironment.

a, Top, mean tumour volume of subcutaneous (4T1) or intradermal (B16, B16-GMCSF) implants in anti-PD-1-, anti-CTLA4- or control-treated mice ($n = 10$). Bottom, survival of 4T1, B16 or B16-GMCSF tumour bearing mice treated with anti-PD1 or anti-CTLA4 compared to control (vehicle treated only) ($n = 10$). **b**, Representative flow cytometric analysis and quantification of CD11b⁺, CD8⁺, CD4⁺ and regulatory T cells (T_{reg} cells; CD4⁺Foxp3⁺) cell populations in 4T1, B16, B16-GMCSF tumours at 14 days post implantation ($n = 5$). **c**, Representative flow cytometric analysis and quantification of granzyme B, PD-1 expression on CD8⁺ T cell populations in 4T1, B16, B16-GMCSF tumours at 7 days post implantation ($n = 5$). **d**, *In vitro* suppressive activity of tumour-infiltrating CD11b⁺ cells purified from 4T1, B16 and B16-GMCSF tumour-bearing mice. Representative histograms of CD8⁺ T cell proliferation at corresponding CD8⁺ to CD11b⁺ T cell ratio (left panel) and quantification of CD8⁺ T cell proliferation using carboxyfluorescein succinimidyl ester (CFSE) dilution (right panel) ($n = 3$). Data represent analysis of n (shown above for each experiment) mice per group, mean \pm s.e.m. Experiment were repeated at least twice to observe concordant statistical significance defined by P value: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ (non-parametric Mann-Whitney U -test, log-rank (Mantel-Cox) test survival comparison).

mice or when IPI-549 is added to the media. A similar functional observation was made in human myeloid suppressor cells in PBMCs (Extended Data Fig. 4b) confirming its potential clinical use. In addition, pharmacodynamic evaluation of whole blood confirms the inhibition of PI3K γ in monocytes after activation with a PI3K γ stimulus in human volunteers (Extended Data Fig. 4a). To confirm that inhibition of PI3K γ in myeloid cells is required to delay tumour progression, we depleted myeloid cells (anti-CD11b) before implanting LLC-Bre tumours in mice. Treatment with IPI-549 did not delay tumour growth in the absence of TAMCs (Extended Data Fig. 2e). Taken together these findings suggest that PI3K γ inhibition using IPI-549 is mainly effective in a tumour landscape rich in suppressive myeloid cells and allows for more precise delineation of patients in which it will potentially yield the greatest activity.

The IPI-549-driven polarization of myeloid cells to a less immunosuppressive phenotype correlates with increased CD8⁺ T cell infiltrates and a higher CD8⁺/regulatory T cell ratio at day 14 in both 4T1 and B16-GMCSF tumours (Fig. 3a). CD8⁺ T cells expressed more granzyme B and proliferated more (Ki67⁺) in IPI-549 treated groups (Fig. 3a and Extended Data Fig. 5a). Along with these effector mechanisms, the expression of PD-1 and CTLA4 were enhanced in T cells (Fig. 3a and Extended Data Fig. 5a). Furthermore, in mice lacking T cells (RAG KO, Nu/Nu or after anti-CD8⁺ T cell depletion) the impact of IPI-549 was abrogated (Fig. 3b and Extended Data Fig. 5b, c). These findings confirm that targeting PI3K γ enhances the T-cell-mediated anti-tumour activity by modifying the suppressive function of myeloid cells. We were concerned that targeting a key pathway in myeloid cells may affect their antigen presentation. Reassuringly, we found that IPI-549 did not reduce, but rather enhanced, the activation of tumour-antigen-specific T cells. This was demonstrated by the transfer of tumour-antigen-specific cells (Pmel-1) in B16-GMCSF tumour-bearing mice (Fig. 3c) and by ELISpot using cells isolated from CT26 tumour-bearing mice (Extended Data Fig. 5d). Additionally, the proportion of effector memory cells (CD44⁺CD62L⁺) in CD8⁺ and CD4⁺ populations was not affected (Extended Data Fig. 6), suggesting that IPI-549 does not affect the ability of antigen-presenting cells to activate antigen-specific T cells. Taken together, these results demonstrate that selective PI3K γ inhibition reduces the immune-suppressive function of myeloid cells, ultimately promoting T effector activation and T-cell-mediated cytotoxicity.

We found that treatment of myeloid-enriched tumours with IPI-549 leads to upregulation of PD-1 and CTLA4 expression on CD8⁺ T cells (Fig. 3a and Extended Data Figs 2d and 5a). This strongly suggests that

constitute only a portion of suppressive TAMCs, we further subdivided myeloid cells into granulocytes (CD11b⁺Ly6G⁺), monocytes (Ly6C^{high}/MHCII^{low}, Mono-Lo), immature macrophages (Ly6C^{high}/MHCII^{high}, Mono-Hi), M1 macrophages (Ly6C^{low}/MHCII^{high}, TAM-M1) or M2 macrophages (Ly6C^{low}/MHCII^{low}, TAM-M2)^{20,21}. When we analysed the myeloid cells with the above gating strategy in 4T1 TILs, we also observed that IPI-549 shifts myeloid cells towards the TAM-M1 population (Extended Data Fig. 3a). Moreover, relative mRNA expression of M1 and M2 markers correlate with a less suppressive function of these cells (Extended Data Fig. 3b). We subsequently tested the suppressive function of myeloid cells derived from IPI-549-treated B16-GMCSF tumour-bearing mice on naive CD8⁺ T cell proliferation (Fig. 2e). We show that suppression of CD8⁺ T cells is abolished in IPI-549-treated

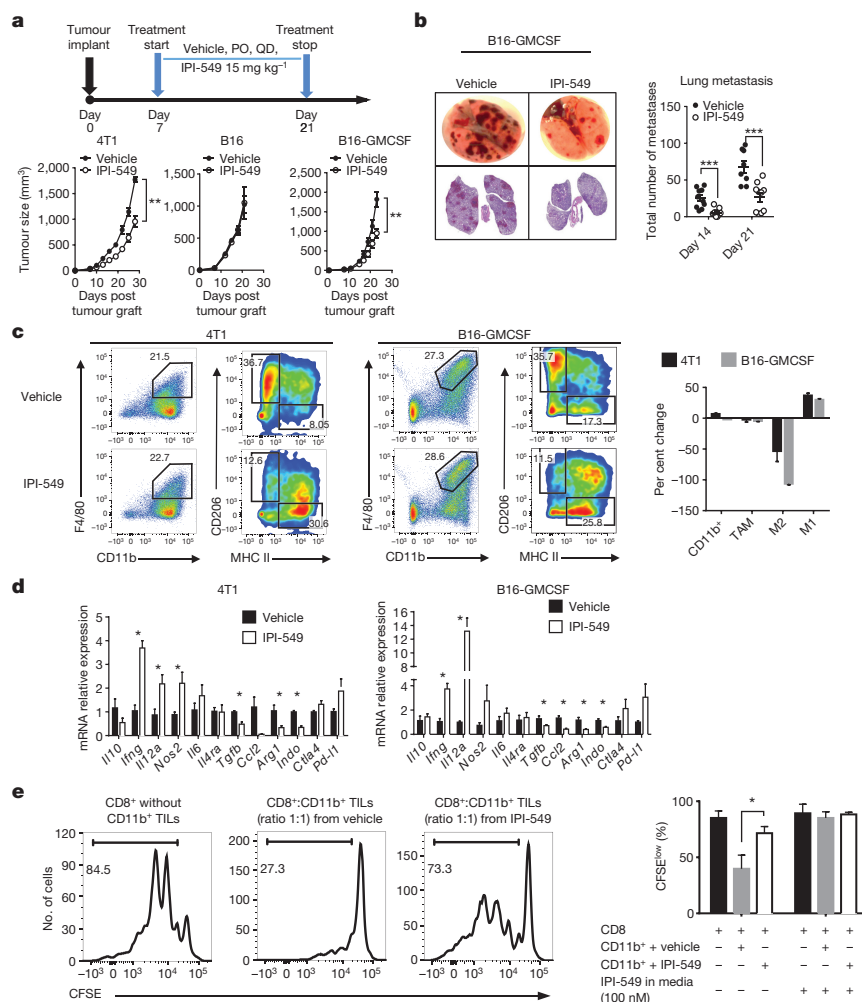


Figure 2 | Selective targeting of PI3K γ reduces tumour growth and metastasis in various checkpoint blockade-resistant tumour models associated with high levels of myeloid cell infiltrates. **a**, Therapy regimen (top panel, PO, *per os*, QD, once a day) and mean tumour volume of subcutaneous (4T1) or intradermal (B16-GMCSF) implants in mice treated with vehicle or IPI-549 (15 mg kg⁻¹ administered orally, daily) ($n = 10$; bottom panel). **b**, Representative pictures and haematoxylin and eosin-stained sections of lung from B16-GMCSF tumour-bearing mice at day 7 on treatment with IPI-549 or vehicle (left panels). Quantification of lung metastasis at day 14 and 21 from tumour challenge (right panel). **c**, Flow cytometric analysis and quantification of CD11b⁺F4/80⁺ (TAM) cell populations in 4T1 tumours at day 7 post implantation ($n = 5$), expression of CD206 (M2) and MHCII (M1) in CD11b⁺F4/80⁺ cell populations. The histogram bars show the percentage change in each cell population (CD11b⁺, TAM, M2 and M1) in IPI-549-treated group in comparison to the vehicle-treated controls. **d**, mRNA expression of selected M1 and M2 markers in 4T1 and B16-GMCSF tumours as determined by RT-PCR ($n = 3$), data were relative to GAPDH expression and normalized versus the mean of vehicle-treated tumours. **e**, *In vitro* suppressive activity of tumour-infiltrating CD11b⁺ cells purified from 4T1 tumour-bearing mice at day 7 post treatment with IPI-549 or vehicle. Representative histograms of CD8⁺ T cell proliferation at a ratio of 1:1 CD8⁺ to CD11b⁺ T cells (left panel) and per cent CD8⁺ T cell proliferation (right panel) ($n = 3$ biological replicates). Data represent analyses of n (shown above for each experiment) mice per group, mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (non-parametric Mann-Whitney U -test).

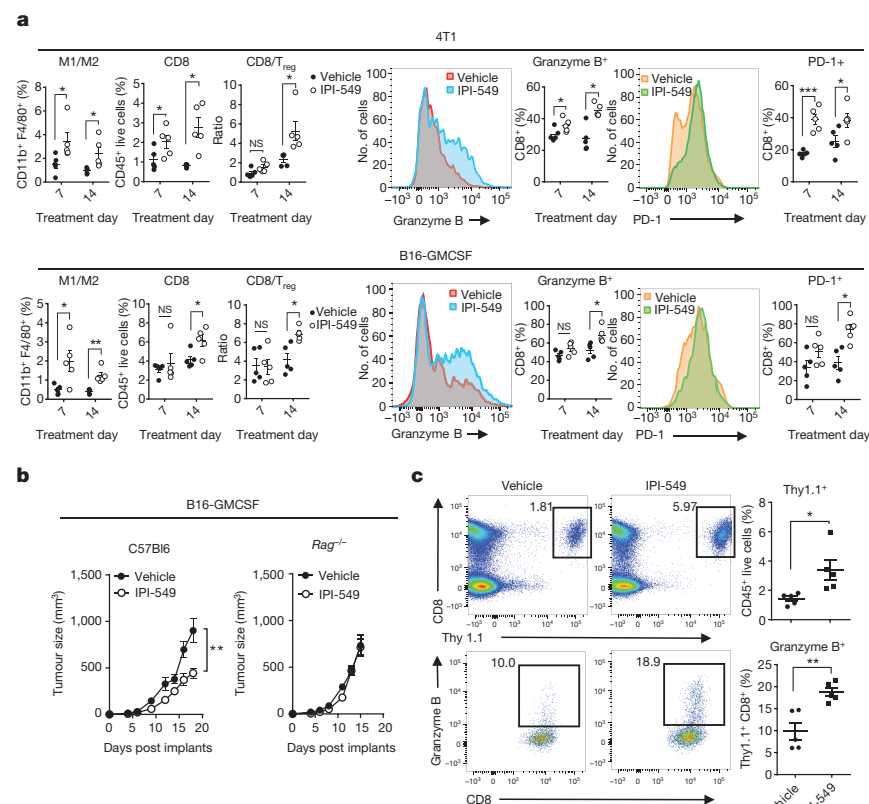


Figure 3 | Reduction of myeloid suppressive phenotype correlates with higher anti-tumour T cell activity. **a**, Left, quantification by flow cytometry of ratio of M1/M2 in CD11b⁺F4/80⁺ cell population, CD8⁺ T cells in CD45⁺ TILs and ratio of CD8⁺/Treg (CD4⁺FoxP3⁺) in CD45⁺ TILs in 4T1 and B16-GMCSF tumours at day 7 and 14 on treatment ($n = 5$). Right, quantification of granzyme B and PD1 expression in CD8⁺ TILs in 4T1 and B16-GMCSF tumours at day 7 and 14 on treatment. **b**, Mean tumour volume of intradermal B16-GMCSF tumours in IPI-549- versus vehicle-treated wild-type C57Bl6 (left) or *Rag*^{-/-} mice (right) ($n = 10$). **c**, Flow cytometric analysis and quantification of gp-100 antigen-specific T cells transferred into IPI-549- or vehicle-treated B16-GMCSF tumour-bearing mice ($n = 5$). Data represent analysis of n (shown above for each experiment) mice per group, mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ (non-parametric Mann-Whitney U -test).

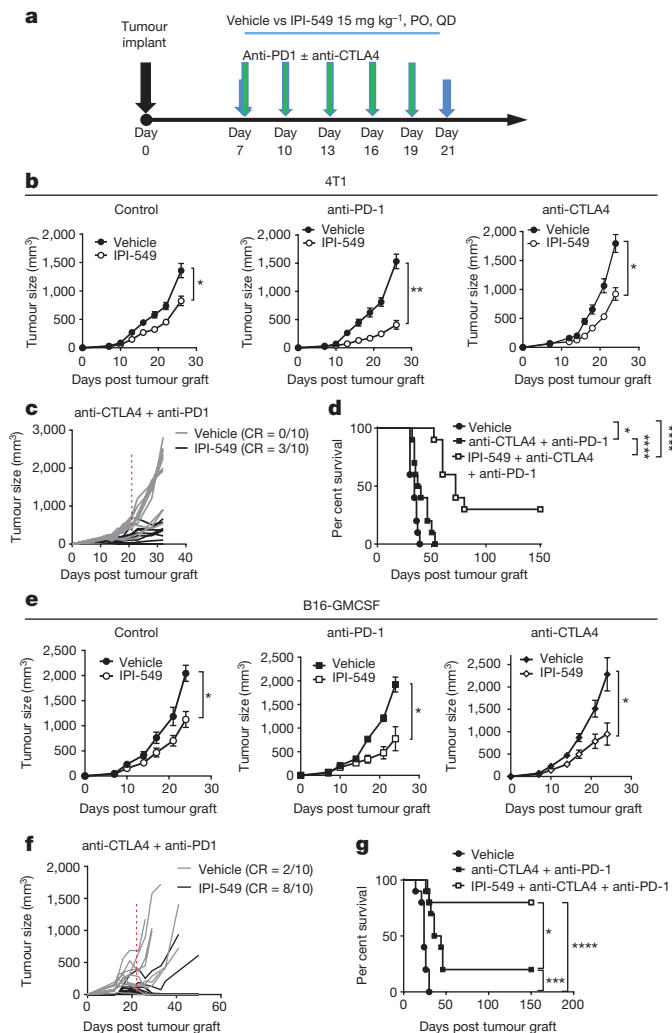


Figure 4 | Resistance to checkpoint blockade therapy is overcome when combined with selective PI3K γ inhibition. **a**, Therapy regimen. **b**, Mean tumour volume of subcutaneous 4T1 tumour in control, anti-CTLA4- or anti-PD-1-treated mice in combination with IPI-549 or vehicle ($n = 10$). **c**, Individual tumour volumes of subcutaneous 4T1 implants in mice treated with anti-CTLA4 and anti-PD1 in combination with IPI-549 or vehicle ($n = 10$), dotted red line corresponds to end of treatment post tumour implantation (day 21). CR, complete response defined as no visible or palpable tumour on consecutive measurements. **d**, Survival at 150 days of 4T1 tumour-bearing mice treated with anti-CTLA4 and anti-PD1 in combination with IPI-549 or vehicle compared to control (vehicle-treated only) ($n = 10$). **e**, Mean tumour volume of intradermal B16-GMCSF implants in anti-PD-1-, anti-CTLA4- or control-treated mice in combination with IPI-549 or vehicle ($n = 10$). **f**, Individual tumour volumes of intradermal B16-GMCSF implants in mice treated with anti-CTLA4 and anti-PD1 in combination with IPI-549 or vehicle ($n = 10$). **g**, Survival at 150 days of B16-GMCSF tumour-bearing mice treated with anti-CTLA4 and anti-PD1 in combination with IPI-549 or vehicle compared to control (vehicle-treated only) ($n = 10$). Data represent analyses of n (shown above for each experiment) mice per group, mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ (non-parametric Mann–Whitney U -test or log-rank test).

combining PI3K γ inhibition with ICB therapy could yield additional anti-tumour activity. In fact, in the ICB-resistant 4T1 or B16-GMCSF models, the combination of either anti-CTLA4 or anti-PD-1 with IPI-549 significantly delayed tumour growth when compared to ICB alone (Fig. 4a–g). We further confirmed this observation in the ICB-resistant LLC tumour model (Extended Data Fig. 7). Analysis of TILs in the combination regimen of IPI-549 with PD-1 blockade shows that the

addition of IPI-549 reduces immune suppression by increasing the M1/M2 ratio, resulting in improved T cell effector function in 4T1 and B16-GMCSF tumours (Extended Data Fig. 8). These results are similar to those reported by Kaneda *et al.* in which PD-1 blockade was combined with a dual PI3K inhibitor (γ/δ)¹⁵; however, given the important effect of PI3K δ in T cell function and more specifically regulatory T cells, it is unclear whether the observed anti-tumour drug effect in that study was myeloid- and/or T-cell-dependent. As CTLA4 and PD-1 have distinct immunologic mechanisms of action, greater therapeutic efficacy is achieved when both therapies are combined²². We tested the anti-PD-1 and anti-CTLA4 combination with IPI-549 in the 4T1 and B16-GMCSF models (Fig. 4a). Double checkpoint blockade therapy alone did not result in any complete tumour regressions in 4T1 tumour-bearing mice and only 20% of the mice bearing B16-GMCSF tumours benefited from the therapy, further confirming multiple checkpoint resistance in these models. Notably, the addition of IPI-549 to the combination of anti-CTLA4 and anti-PD-1 was associated with complete remissions in 30% of 4T1- and 80% of B16-GMCSF tumour-bearing mice. Importantly, tumour-free survivors were resistant to tumour re-implantation (Extended Data Fig. 9), indicating long-lasting adaptive immunity.

Taken together, our data indicate that high infiltration by suppressive TAMCs promotes resistance to ICB, further confirming the complex, multi-factorial mechanisms of immune suppression in the tumour microenvironment. We show that modulating the suppressive phenotype of these cells towards a more inflammatory one can be achieved by selective pharmacologic targeting of the myeloid PI3K γ isoform. By doing so, sensitivity to ICB can be effectively restored in TAMC-rich tumours. These findings provide a strong rationale to consider exploring PI3K γ inhibition to overcome resistance to ICB in clinical trials based on a precision-medicine-type assessment of the tumour immune landscape. This hypothesis is currently being tested in the IPI-549 Phase I/Ib trial, as both a monotherapy and in combination with PD-1 blockade in solid tumours (clinicaltrials.gov identifier NCT02637531).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 September; accepted 24 October 2016.

Published online 9 November 2016.

- Khalil, D. N., Smith, E. L., Brentjens, R. J. & Wolchok, J. D. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nat. Rev. Clin. Oncol.* **13**, 273–290 (2016).
- Coussens, L. M., Zitvogel, L. & Palucka, A. K. Neutralizing tumor-promoting chronic inflammation: a magic bullet? *Science* **339**, 286–291 (2013).
- Gajewski, T. F., Schreiber, H. & Fu, Y.-X. Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* **14**, 1014–1022 (2013).
- Schmid, M. C. & Varner, J. A. Myeloid cells in the tumor microenvironment: modulation of tumor angiogenesis and tumor inflammation. *J. Oncol.* **2010**, 201026 (2010).
- Diaz-Montero, C. M., Finke, J. & Montero, A. J. Myeloid-derived suppressor cells in cancer: therapeutic, predictive, and prognostic implications. *Semin. Oncol.* **41**, 174–184 (2014).
- Gebhardt, C. *et al.* Myeloid cells and related chronic inflammatory factors as novel predictive markers in melanoma treatment with Ipilimumab. *Clin. Cancer Res.* **21**, 5453–5459 (2015).
- De Vlaeminck, Y., González-Rascón, A., Goyvaerts, C. & Breckpot, K. Cancer-associated myeloid regulatory cells. *Front. Immunol.* **7**, 113 (2016).
- Highfill, S. L. *et al.* Disruption of CXCR2-mediated MDSC tumor trafficking enhances anti-PD1 efficacy. *Sci. Transl. Med.* **6**, 237ra67 (2014).
- Meyer, C. *et al.* Frequencies of circulating MDSC correlate with clinical outcome of melanoma patients treated with ipilimumab. *Cancer Immunol. Immunother.* **63**, 247–257 (2014).
- Bjoern, J. *et al.* Immunological correlates of treatment and response in stage IV malignant melanoma patients treated with Ipilimumab. *Oncol. Immunology* **5**, e1100788 (2015).
- Holmgaard, R. B., Zamarin, D., Lesokhin, A., Merghoub, T. & Wolchok, J. D. Targeting myeloid-derived suppressor cells with colony stimulating factor-1 receptor blockade can reverse immune resistance to immunotherapy in indoleamine 2,3-dioxygenase-expressing tumors. *EBioMedicine* **6**, 50–58 (2016).

12. Ugel, S., De Sanctis, F., Mandruzzato, S. & Bronte, V. Tumor-induced myeloid deviation: when myeloid-derived suppressor cells meet tumor-associated macrophages. *J. Clin. Invest.* **125**, 3365–3376 (2015).
13. Gabrilovich, D. Mechanisms and functional significance of tumour-induced dendritic-cell defects. *Nat. Rev. Immunol.* **4**, 941–952 (2004).
14. Lesokhin, A. M. *et al.* Monocytic CCR2⁺ myeloid-derived suppressor cells promote immune escape by limiting activated CD8 T-cell infiltration into the tumor microenvironment. *Cancer Res.* **72**, 876–886 (2012).
15. Kaneda, M. M. *et al.* PI3K γ is a molecular switch that controls immune suppression. *Nature* <http://dx.doi.org/10.1038/nature19834> (2016).
16. Hirsch, E. *et al.* Central role for G protein-coupled phosphoinositide 3-kinase γ in inflammation. *Science* **287**, 1049–1053 (2000).
17. Li, Z. Roles of PLC- β 2 and - β 3 and PI3K γ in chemoattractant-mediated signal transduction. *Science* **287**, 1046–1049 (2000).
18. Schmid, M. C. *et al.* Receptor tyrosine kinases and TLR/IL1Rs unexpectedly activate myeloid cell PI3K γ , a single convergent point promoting tumor inflammation and progression. *Cancer Cell* **19**, 715–727 (2011).
19. Evans, C. A. *et al.* Discovery of a selective phosphoinositide-3-kinase (PI3K) γ inhibitor (IPI-549) as an immuno-oncology clinical candidate. *ACS Med. Chem. Lett.* **7**, 862–867 (2016).
20. Laoui, D. *et al.* Tumor hypoxia does not drive differentiation of tumor-associated macrophages but rather fine-tunes the M2-like macrophage population. *Cancer Res.* **74**, 24–30 (2014).
21. Movahedi, K. *et al.* Different tumor microenvironments contain functionally distinct subsets of macrophages derived from Ly6C(high) monocytes. *Cancer Res.* **70**, 5728–5739 (2010).
22. Larkin, J. *et al.* Combined Nivolumab and Ipilimumab or monotherapy in untreated melanoma. *N. Engl. J. Med.* **373**, 23–34 (2015).

Acknowledgements We would like to thank the Flow Cytometry and Integrated Genomics Operation Core Facilities at MSKCC. Swim Across America, Ludwig Institute for Cancer Research, Parker Institute for Cancer Immunotherapy, Center for Experimental Therapeutics at MSKCC (ETC), and the Breast Cancer Research Foundation supported this work. The work was also supported in part by the MSKCC Core Grant (P30 CA008748). O.D.H. was supported by J. Houtard foundation, Nuovo Soldati Foundation and Wallonie-Bruxelles International. We would also like to thank Y. Senbabaoglu for his help in bioinformatics data analysis, A. Bossert for his contribution as part of the GME program as well as J. Gladstone and K. Walsh for their contributions while working as co-op students in the laboratory.

Author Contributions O.D.H., T.M., J.D.W., K.M., J.L.K., V.J.P. and J.A. developed the concepts and discussed experiments. O.D.H., T.M., J.D.W., K.M. and J.L.K. wrote the manuscript. O.D.H., M.R., D.W., L.F.C., D.H.C., S.B., A.G., M.P., J.P. and N.K. performed and analysed animal model experiments, flow cytometry experiments and functional assays. C.L. provided technical assistance; S.S. and K.W. performed assays in human samples. M.D., T.T. and H.S. performed transcriptomic analysis. J.T. and J.S. performed pharmacodynamics and pharmacokinetics studies.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M. (merghout@mskcc.org) or J.D.W. (wolchokj@mskcc.org).

Reviewer Information *Nature* thanks F. Balkwill, M. De Palma and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cell lines. The murine cancer cell lines for melanoma (B16F10, referred to as B16), breast cancer (4T1) and colon cancer (CT26) were obtained from ATCC. The colon cancer cell line MC38 was obtained from the NCI and the lung cancer (LLC-Brei) from Caliper Life Sciences. Cells were maintained in RPMI medium supplemented with 10% fetal calf serum (FCS) and penicillin with streptomycin (complete RPMI media). The GM-CSF-secreting B16 cell line (referred to as B16-GM-CSF) has been reported and was used to increase the number of myeloid cells recruited to the tumour¹⁴. The cell lines have been mycoplasma tested.

Quantitative RT-PCR. cDNA was prepared using 1 µg RNA with the qScript cDNA Synthesis Kit (Quanta Biosciences). Sybr green-based qPCR was performed using murine primers to *Arg1*, *Ifng*, *Il10*, *Il12p40*, *Il1b*, *Il6*, *Ccl2*, *Gapdh*, *Nos2*, *Tgfb1*, *Tnfa*, *Il4ra*, *Indo*, *Ctla4*, *Pd-1*, *CD86*, *Cxcr2*, *Fizz1*, *Ymd1* (Qiagen QuantiTect Primer Assay). mRNA levels were normalized to *Gapdh* ($\Delta C_t = C_t$ gene of interest $- C_t$ *Gapdh*) and reported as relative mRNA expression ($\Delta\Delta C_t = 2^{-(\Delta C_{t(\text{sample})} - \Delta C_{t(\text{control})})}$) or fold change.

Cytokine analysis. Tumours were excised, snap-frozen in liquid nitrogen, and pulverized using a tissue grinder. Tumour protein lysates were prepared in MSD Tris Lysis Buffer (Meso Scale Discovery, Rockville, Maryland) containing 2× Halt Protease and Phosphatase Inhibitor Cocktail (Fisher Scientific). Total protein concentration was normalized to 4 mg ml⁻¹ and cytokines were quantified using the MILLIPLEX MAP Mouse Cytokine/Chemokine Magnetic Bead 32 Plex Panel and MILLIPLEX MAP Mouse CD8⁺ T Cell Magnetic Bead Panel kits according to the manufacturer's instructions (Millipore, Billerica, Massachusetts).

RNA-seq. Mice bearing CT26 tumours were treated with vehicle or IPI-549 (15 mg kg⁻¹ day⁻¹, PO) for 6 or 9 days. Tumours were isolated, and frozen until needed; tumours then thawed and RNA was extracted from all cells. RNA-seq was done at Expression Analysis (Q2 Solutions). Sequence reads were aligned to the mouse B38 reference genome using OmicSoft ArrayStudio and the UCSC gene model. log₂[FPKM] was calculated for each gene, and data were mean centred for display in heat maps. The analysis focused on a compilation of about 4,200 mouse genes related to cancer immunology and PI3K pathway signalling compiled from numerous sources including BioCarta pathways, GO gene ontologies, KEGG pathways, WikiPathways, and literature²³.

Mice. Female C57BL/6J and Balb/c mice (6–8 weeks old) were purchased from Jackson Laboratory. Pmel-1 TCR transgenic mice have been previously reported²⁴ and were provided by N. Restifo (National Cancer Institute, Bethesda, Maryland). All mice were maintained in micro-isolator cages and treated in accordance with the NIH and American Association of Laboratory Animal Care regulations. All mouse procedures and experiments for this study were approved by the Memorial Sloan Kettering Cancer Center Institutional Animal Care and Use Committee. No statistical methods were used to predetermine sample size. Ten to fifteen mice per treatment strategy were used to allow 90% power, and a 5% significance level, and detect differences in tumour-free survival from 10% to 80%. Typically, tumours grew in 100% of control animals. An additional 5 mice per group were used for tissue harvest at day 7 and day 14 on treatment. Mice age and treatment allowance were randomized at day 7 after tumour implants.

Tumour challenge and treatment experiments. On day 0 of the experiments, tumour cells were injected intradermally (i.d.) in the right flank. For the B16 model, 2.5×10^5 B16-WT or B16-GM-CSF cells were injected into C57BL/6J mice. For 4T1 model and for the CT26 model, 5×10^5 cells were used subcutaneously in Balb/c mice. For studies in immune compromised mice, the CT26 study was done in the Balb/c NU/NU strain and the B16-GM-CSF in C57BL/6 *Rag1*^{-/-} mice. Mice were obtained from Jackson Labs and Charles River Labs. Treatments were given as single agents or in combinations with the following regimen for each drug. The PI3Kγ inhibitor drug IPI-549 was dissolved at 5% 1-methyl-2-pyrrolidinone in polyethylene glycol 400 and administered by oral gavage once a day at 15 mg kg⁻¹. Treatment was initiated on day 7 ending on day 21 post tumour implant. Control groups received vehicle (5% NMP, 95% PEG) without the active product. Anti-CTLA4 antibody (100 µg per mouse, clone 9H10, Bio X cell) and anti-PD-1 antibody (250 µg per mouse, clone RPM1-14, Bio X cell) were injected intraperitoneally (i.p.) on days 7, 10, 13 and 16 for the B16, B16-GM-CSF and 4T1 models and on days 10, 13 and 16 for the CT26 model. Tumours were measured every second or third day with a calliper, and the volume (length × width × height) was calculated. Mice that had no visible and palpable tumours that could be measured on consecutive measurement days were considered complete regressions. Animals were euthanized for signs of distress or when the total tumour volume reached 2,500 mm³. For the re-challenge study: mice with complete responses in the anti-PD-1 treatment group and the anti-PD-1 and IPI-549 combination group were re-challenged with 2.5×10^5 CT26 WT tumour cells (on day 106 of the study since original tumour implant). Additional mice with complete responses from an additional IPI-549 and anti-PD-1 group were implanted with 1×10^5 4T1 tumour cells.

Cell depletion studies. Female BALB/c mice were depleted of CD8⁺ T cells by intraperitoneal (IP) injection of a CD8-depleting antibody beginning 3 days before tumour implantation and continuing every 3 days (Q3D) for the duration of the study. Control animals were injected IP with an isotype control antibody according to the same dosing schedule. CD8⁺ T cell depletion was verified by flow cytometry of splenic cells from a separate cohort of mice at the time of tumour implant. Anti-CD8 antibody (BioXCell, *in vivo* MAb Rat IgG2b anti-mouse CD8, clone 169.4) and isotype control antibody (BioXCell, *in vivo* MAb Rat IgG2b, κ isotype control, clone LTF-2) were used. CD8⁺ T-cell-depleted and isotype control-treated mice were implanted with 2.5×10^5 CT26 WT cells subcutaneously to the dorsal right flank. Eight days after implant, tumour-bearing animals with average tumour volumes of 50 to 60 mm³ began treatment. Animals were dosed daily with vehicle or IPI-549 (15 mg kg⁻¹, PO) and dosed every three days with 100 µg antibody, either isotype control or anti-CD8. Tumour measurements were taken every second or third day during the 13 days of dose administration. For CD11b⁺ cell depletion, murine Lewis Lung Carcinoma tumour brei (here referred to as LLC) was propagated into ten C57BL/6 Albino male mice. When tumours reached an average of 1,500 mm³, tissue was collected and made into a single-cell suspension. Tissue was dissociated in a glass dounce, filtered through a 100 µm filter and washed with cold buffer of PBS pH 7.2, 0.5% BSA and 2 mM EDTA. An aliquot of cells was separated, and placed on ice until the time of re-transplant. Cells were counted and adjusted to 2×10^7 total cells. To deplete the CD11b cells from the tumour homogenate, CD11b micro beads (MACS #130-049-601) were added to the sample and incubated on ice for 15 min. Cells were washed and re-suspended with buffer. A negative depletion was performed following the Miltenyi autoMACS protocol. The positive selected cells and starting tumour inoculum were counted and cell numbers adjusted to the tumour inoculum used for historical studies. For each cell condition 30 C57BL/Albino male mice were inoculated. Treatment of IPI-549 or vehicle started when tumours for each cell condition reached an average of 200 mm³ regardless of what day post implant this occurred.

Isolation of tumour-infiltrating cells and lymphoid tissue cells. Mouse tumour samples were minced with scissors before incubation with 1.67 U ml⁻¹ Liberase (Roche) and 0.2 mg ml⁻¹ DNase (Roche) in RPMI for 30 min at 37 °C. Tumour samples were homogenized by repeated pipetting and filtered through a 100-µm nylon filter (BD Biosciences) in RPMI supplemented with 7.5% FCS to generate single-cell suspensions. Cell suspensions were washed once with complete RPMI and purified on a Ficoll gradient to eliminate dead cells. Cells from mouse spleens were isolated by grinding spleens through 40-µm filters. After red blood cell (RBC) lysis (ACK Lysing Buffer, Lonza) when required, all samples were washed and re-suspended in FACS buffer (PBS/0.5% albumin) or RPMI depending on further use.

Flow cytometry and morphology analysis. Cells isolated from mouse tumours and spleens were pre-incubated (15 min, 4 °C) with anti-CD16/32 monoclonal antibody (Fc block, clone 2.4G, BD Biosciences) to block nonspecific binding and then stained (30 min, 4 °C) with appropriate dilutions of various combinations of the following fluorochrome-conjugated antibodies: anti-CD45-AF 700 (clone 30-F11), anti-CD11b-APC-Cy7 (clone M1/70), anti-CD11b-PE-TR(M1/70.15), anti-Ly6G-APC (clone 1A8), anti-F4/80-PerCP-Cy5.5 (clone BM8), anti-Ly6C-PE-Cy7 (clone AL-21), anti-MHC Class II-eFluor 450 (clone M5/114.15.2), anti-CD206-PE (clone 19.2), anti-CD8-PerCP-Cy5.5 (clone 53-6.7), anti-CD8-PE Texas Red (clone 5H10), anti-CD4-PE-Cy7 (clone RM4-5), anti-CD4-Pacific Blue (clone RM4-5), anti-Foxp3-APC (clone FJK-16s), anti-Foxp3-PE-Cy7 (clone FJK-16s), anti-CD25-APC-Cy7 (clone PC61), anti-CD44-PE-Cy7 (clone IM7), anti-CD62L-PE (clone MEL-14), anti-Ki67 (clone B56), anti-CTLA4-APC (clone), anti-Granzyme B-PE-TR (clone GB11), antibodies, all purchased from BD Biosciences, eBioscience or Invitrogen. For intracellular stain, cells were further permeabilized using a FoxP3 Fixation and Permeabilization Kit (eBioscience) and stained for Foxp3 (clone FJK-16s, Alexa-Fluor-700-conjugated, eBioscience), Ki67 (clone SolA15, eFluor-450-conjugated, eBioscience) or CTLA4. The stained cells were acquired on a LSRII Flow Cytometer using BD FACSDiva software (BD Biosciences) and the data were processed using FlowJo software (Treestar). Dead cells and doublets were excluded on the basis of forward and side scatter and Fixable Viability Dye eFluor 506.

Purification of myeloid or CD8⁺ T cells from tumours or spleen. Mouse tumour and spleen single-cell suspensions were generated as described in the previous section. Tumour cells were subsequently separated from debris over a Ficoll gradient (Sigma-Aldrich). B cells were depleted from splenocytes using CD19 microbeads and LD columns according to the manufacturer's instructions (Miltenyi Biotec) to enrich the myeloid fractions. Cells were stained with anti-CD45.2-Alexa-Fluor-700, anti-CD11b-APC-Cy7, anti CD8-PE antibodies for flow sorting on a FACSaria II Cell Sorter (BD Biosciences). Dead cells were excluded using DAPI (Invitrogen). Purity of flow-sorted populations was above 90%.

Isolation of Pmel lymphocytes and adoptive transfer. Spleens and lymph nodes from *pmel-1* TCR transgenic mice were isolated and ground through 100- μ m filters. After RBC lysis, CD8⁺ T cells were purified by positive selection using Miltenyi magnetic beads. The isolated cells were loaded with CellTrace Far Red DDAO-s.e. (Thermo Fisher Scientific) and injected into recipient animals via tail vein at indicated numbers. Activated CD8⁺ T cells were generated by culturing splenocytes with soluble anti-CD3 (1 μ g ml⁻¹, clone 145-2C11, eBioscience) and anti-CD28 (2 μ g ml⁻¹, clone 37.51, eBioscience) for 72 h. Recombinant mouse IL-2 (30 U ml⁻¹, Chiron) was added for the final 24 h of culture. CD8⁺ T cells were subsequently positively selected with Miltenyi magnetic beads before injection via tail vein, as described above. The frequency and proliferation of Pmel cells were measured 2 weeks after tumour challenge and 7 days after adoptive transfer of 1×10^6 *in vitro* activated CD8⁺ Pmel T cells using Thy1.1 antibody and by assessing CellTrace Far Red DDAO-s.e. dilution by flow cytometry, respectively.

T cell suppression assay. Spleens from naive mice were isolated and ground through 40- μ m filters to generate a single-cell suspension. After RBC lysis, CD8⁺ cells were purified using anti-CD8 (Ly-2) microbeads (Miltenyi Biotec) according to the manufacturer's protocol and labelled with 1 mM CFSE (Invitrogen) in pre-warmed PBS for 10 min at 37°C. The CFSE-labelled CD8⁺ T cells were then plated in complete RPMI media supplemented with 0.05 M β -mercaptoethanol onto round bottom 96-well plates (25×10^3 cells per well) coated with 1 μ g ml⁻¹ anti-CD3 (clone 1454-2C11) and 5 μ g ml⁻¹ anti-CD28 (clone 37N) antibodies. Purified myeloid cells were added in indicated ratios and plates were incubated at 37°C. After 48 h, cells were harvested and CFSE signal in the gated CD8⁺ T cells was measured by flow cytometry (LSRII Flow Cytometer, BD Biosciences). For the human MDSC suppression assay PBMCs were isolated using Lymphoprep from donor blood. T cells were isolated by CD3⁺ selection (Easysep, Stem Cell Technologies) and frozen in Sigma freezing media for later use. The remaining PBMCs without T cells were incubated with 20 ng ml⁻¹ GM-CSF and 20 ng ml⁻¹ IL-6 for 6 days to differentiate the myeloid cells and were incubated with or without added IPI-549. After 6 days, the MDSC cells were isolated by CD33⁺ selection (Easysep, Stem Cell Technologies). These MDSC cells were then mixed with the autologous T cells (at a 1:4 ratio) that had been pre-stained with cell trace violet and activated with anti-CD3 and anti-CD28 beads (Dyna). The T cell proliferation is determined by Cell Trace Violet dye dilution measured by flow cytometry after 72 h.

IFN γ ELISPOT assay. Blood was collected from CT26 tumour-bearing mice after 10 days of treatment with either vehicle or IPI-549 (15 mg kg⁻¹). PBMCs were isolated using Lymphoprep density gradient media (Stem Cell Technologies, Vancouver, BC). IFN γ producing cells were quantified using the CTL Mouse Immunospot IFN γ Single Colour ELISPOT kit (CTL, Shaker Heights, Ohio) according to the manufacturer's instructions. For *in vitro* re-stimulation, 1×10^5 PBMC were co-cultured with 1×10^5 irradiated (2,000 rad) CT26 colon carcinoma target cells in CTL test media (CTL) for 16 h. Irradiated (2,000 rad) 4T1 mammary

carcinoma target cells were used as negative control targets to assess specificity. IFN γ spots were quantified using a CTL Immunospot S6 Micro Analyzer and Immunospot Professional Software (CTL).

Macrophage polarization assay. Bone marrow derived macrophages were prepared from C57BL/6 mice femur and tibias. Red blood cells were lysed and then the remaining cells were plated in bone macrophage media (BMM) consisting of DMEM, 20% FBS plus penicillin/streptomycin and 50 ng ml⁻¹ M-CSF and incubated for 6 days. Cells were polarized towards an M2 phenotype with the addition of 20 ng ml⁻¹ IL4 and 50 ng ml⁻¹ M-MCSF (both from R&D Systems) with or without added IPI-549. Cells were incubated for 48 h and then RNA was harvested from the cells (Qiagen RNeasy). qRT-PCR was performed using primers for mouse ARG1 (Mm00475988_m1, Life Technologies NY) and mouse β -actin (Mm00607939_s1 Life Technologies, New York).

Biochemical and cell-based assays. Biochemical and cell-based assays for the Class I and Class II PI3K isoforms were run as previously described^{19,25}.

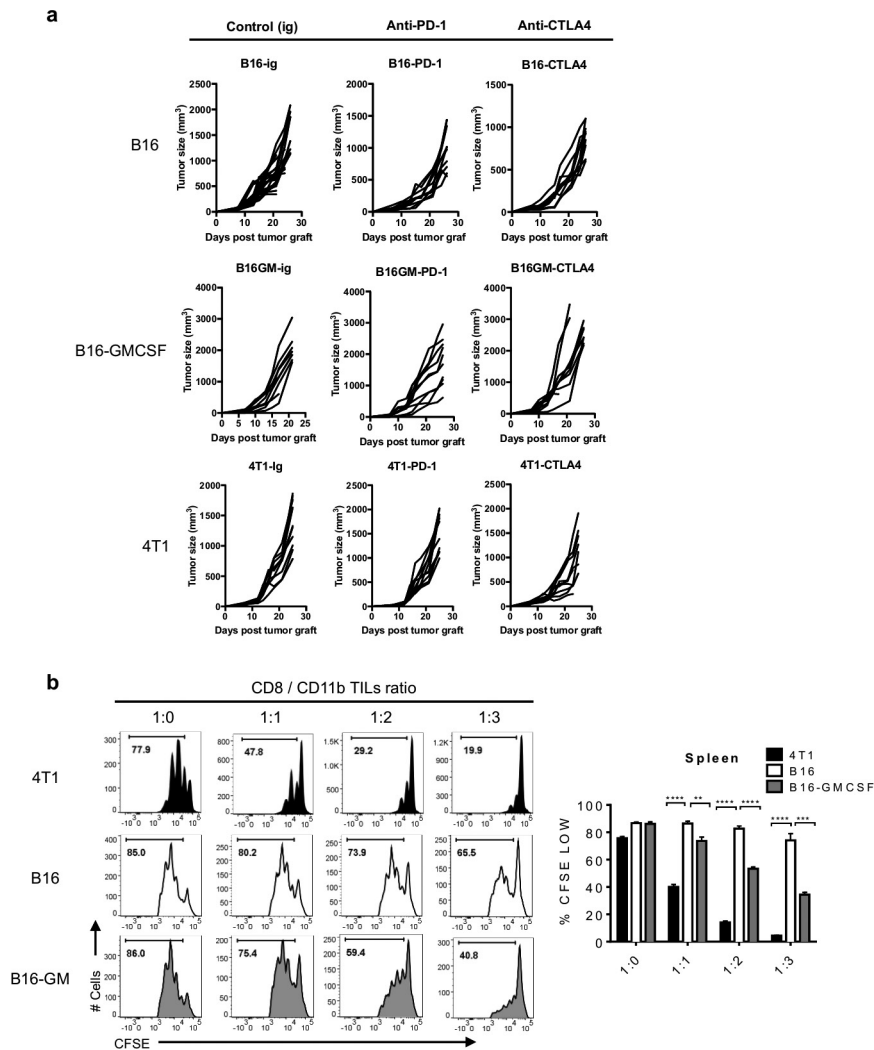
PI3K γ -specific whole-blood PD assay. Whole blood from six healthy donors was pre-treated with a dose titration of IPI-549 and then stimulated for 2 min with 2.3 μ g ml⁻¹ CXCL12. Cells were lysed, fixed and stained. Human blood samples were collected after ICF approval. Response to stimulation was determined by measuring phosphorylation of AKT S473 in monocytes (CD14⁺) by flow cytometry and comparing the value to that of untreated controls. IC₅₀ values for IPI-549 were calculated for each donor, measuring compound potency against PI3K γ and in whole blood.

Statistics. Where indicated, data were analysed for statistical significance and reported as *P* values. Data were analysed with a Mann–Whitney test when comparing means of two independent groups and two-way ANOVA when comparing more than two groups. *P* < 0.05 was considered statistically significant (**P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001). Evaluation of survival patterns in tumour-bearing mice was performed using the Kaplan–Meier method, and results were ranked according to the Mantel–Cox log-rank test. *P* < 0.05 was considered statistically significant. Survival was defined as mice with tumours <2,500 cm³.

The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. Experiments have been repeated to ensure reproducibility of the observations.

Data availability. Source Data for the main figures are provided in the online version of this paper.

23. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).
24. Overwijk, W. W. *et al.* Tumour regression and autoimmunity after reversal of a functionally tolerant state of self-reactive CD8⁺ T cells. *J. Exp. Med.* **198**, 569–580 (2003).
25. Winkler, D. G. *et al.* PI3K- δ and PI3K- γ inhibition by IPI-145 abrogates immune responses and suppresses activity in autoimmune and inflammatory disease models. *Chem. Biol.* **20**, 1364–1374 (2013).

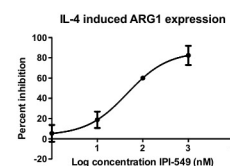


Extended Data Figure 1 | Effect of suppressive myeloid TILs in response to checkpoint blockade. **a**, Individual tumour growth of subcutaneous (4T1) or intradermal (B16, B16-GMCSF) implants in anti-PD-1-, anti-CTLA4- or control-treated mice ($n = 10$). **b**, *In vitro* suppressive activity of tumour-infiltrating CD11b⁺ cells purified from spleen of 4T1,

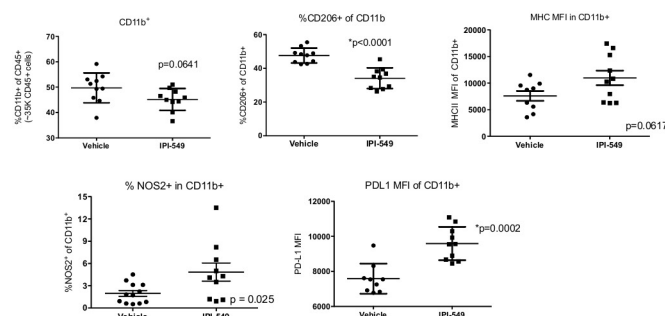
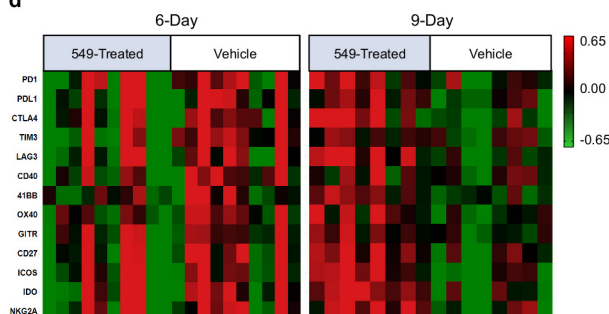
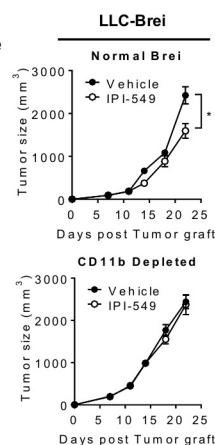
B16, B16-GMCSF tumour-bearing mice. Representative histograms of CD8⁺ T cell proliferation at corresponding CD11b⁺ to CD8⁺ T cell ratio (left panel) and quantification of CD8⁺ T cell proliferation (right panel) ($n = 3$), mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ (non-parametric Mann–Whitney U -test).

a

Assay (Mean values)	PI3K- α	PI3K- β	PI3K- γ	PI3K- δ
Binding affinity K_d = K_{off}/K_{on} (nM) [n]	17 [2]	82 [2]	0.29 [3]	23 [3]
Biochemical IC_{50} (nM) [n]	3241 [4]	3468 [4]	15.7 [4]	> 8421 [4]
Cellular IC_{50} (nM) [n]	247.7 [4]	241.0 [2]	1.2 [2]	175.8 [2]

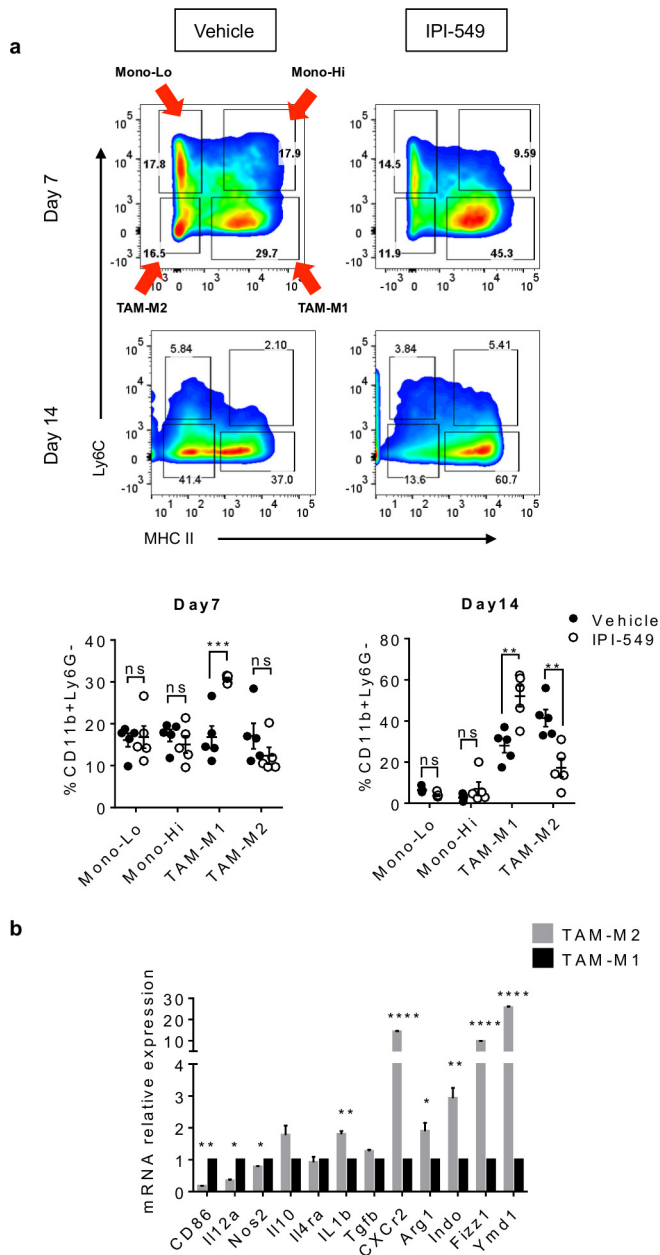
**b**

Cancer type	Cell Line	% Tumor growth inhibition range [n experiments]
Lung	LLC-luc-Brei	28-52% [10]
Colon	MC38	40-45% [3]
Breast	4T1	19-38% [6]
Colon	CT26	24-45% [12]
Melanoma	B16-GMCSF	24-56% [6]

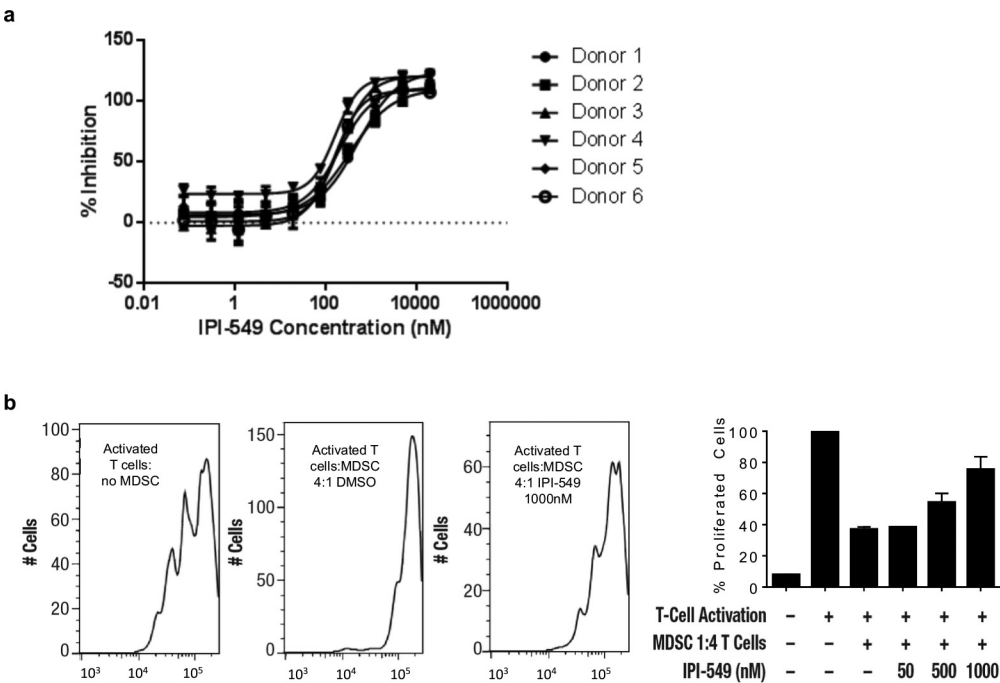
c**d****e**

Extended Data Figure 2 | Effect of selective PI3K γ inhibition on tumour growth and myeloid TILs. **a**, Binding affinities (K_d) and cellular IC_{50} inhibition of pAKT by IPI-549 for class I PI3K isoforms (left table). Percentage of inhibition of expression on bone-marrow-derived macrophages polarized with M-CSF and IL-4, (right panel). **b**, Percentage of tumour growth inhibition in LLC, MC38, 4T1, CT26, B16-GMCSF tumour-bearing mice treated with IPI-549 (table). **c**, Quantification of CD11b, CD206, NOS2 and PD-L1 expression in CD11b $^{+}$ tumour-

infiltrating leukocytes from IPI-549- versus vehicle-treated CT26 tumour-bearing mice. **d**, RNA-seq of co-stimulatory and checkpoint molecules on whole tumours from CT26 tumour-bearing mice treated for 6 or 9 days with IPI-549 compared to vehicle. **e**, Mean tumour volume of subcutaneous LLC-Brei implants in IPI-549- versus vehicle-treated mice without or after CD11b $^{+}$ cell depletion. Data represent analysis of 5–10 mice per group, mean \pm s.e.m. * $P < 0.05$, *** $P < 0.001$ (non-parametric Mann–Whitney U -test).

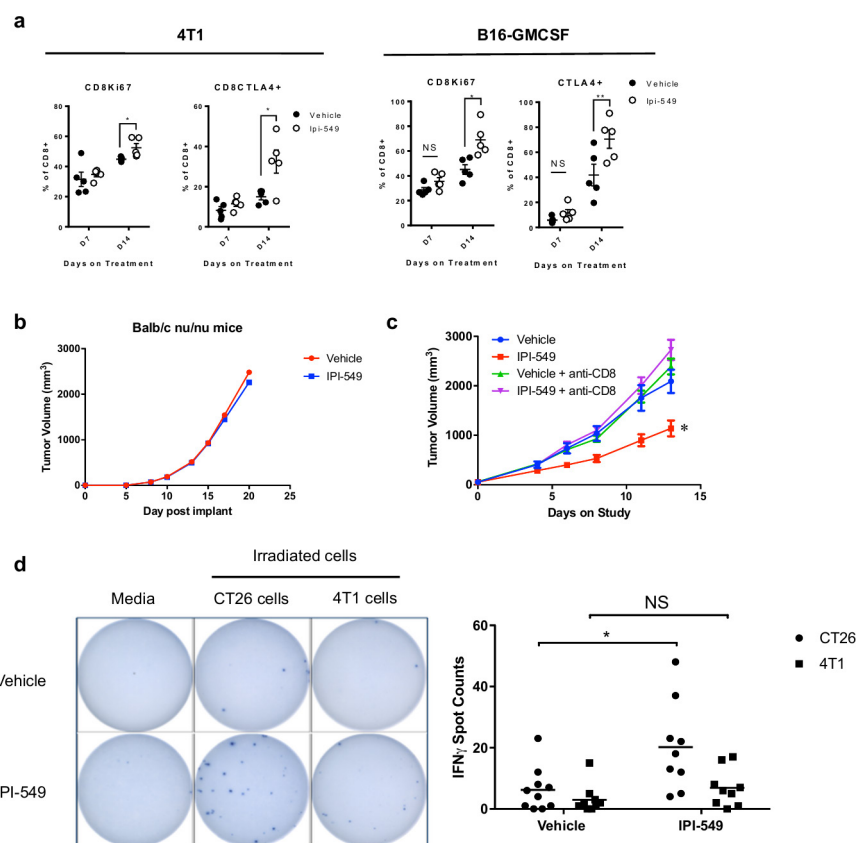


Extended Data Figure 3 | Effect of selective PI3K γ inhibition on subsets of CD11b myeloid cells. **a**, Representative flow cytometry analysis and quantification of Ly6C, MHC class II expression in CD11b⁺Ly6G⁻ cells infiltrating 4T1 tumours. **b**, mRNA expression of selected M1 and M2 markers in sorted Ly6C^{low}MHCII^{low} (TAM-M2) compared to Ly6C^{low}MHCII^{high} (TAM-M1) population from 4T1 tumour, data were relative to GAPDH expression and normalized versus the mean of TAM-M1 population. Mean \pm s.e.m. * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001 (non-parametric Mann-Whitney U -test).



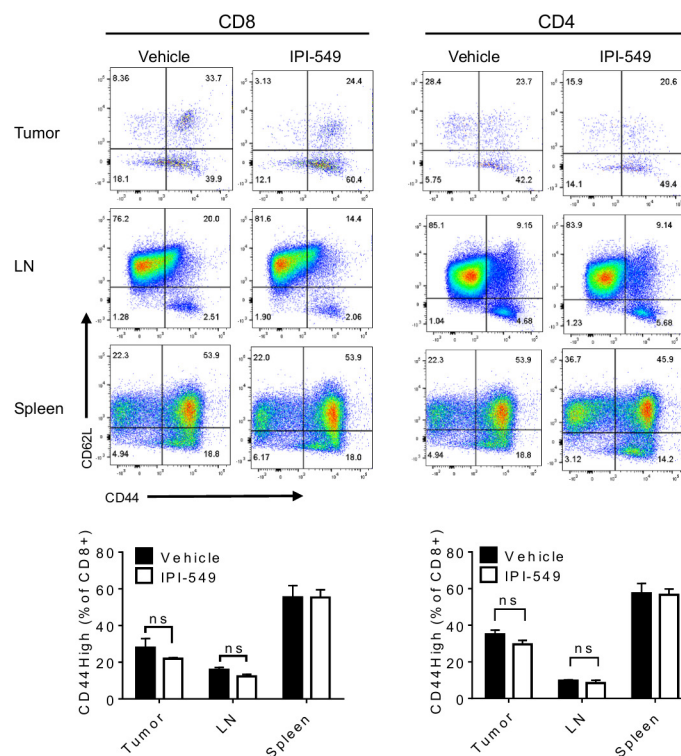
Extended Data Figure 4 | Effect of selective PI3K γ inhibition on suppressive PBMC derived human myeloid cells. **a**, Inhibition of CXCL-12 activation of PI3K γ in monocytes as measured by pAKT (S473) in human whole blood. **b**, Representative histograms and quantification of

human CD8⁺ T cell proliferation after 72 h of co-culture with or without autologous myeloid-derived suppressor cells generated from the T-cell-depleted PBMCs \pm IPI-549.



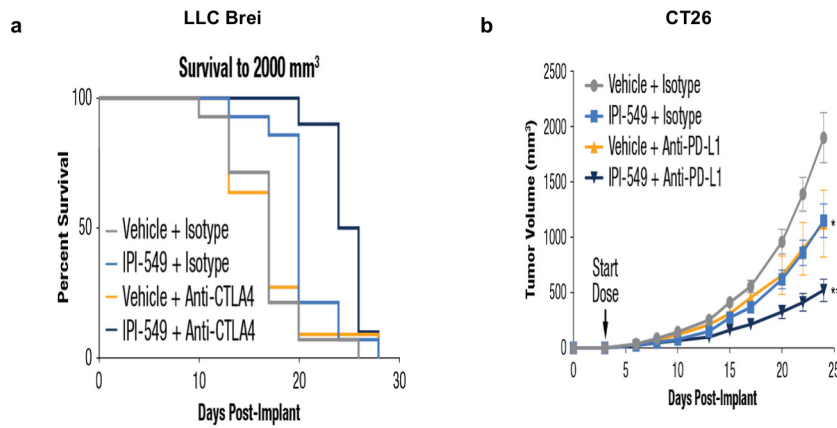
Extended Data Figure 5 | Effect of selective PI3K γ inhibition on function of tumour specific T cell responses. **a**, Quantification of KI67, and CTLA4 expression in CD8⁺ T cells in TILs of 4T1 or B16-GMCSF tumours at IPI-549 compared to vehicle treatment days 7 and 14. **b**, Mean tumour volume of subcutaneous 4T1 tumour in IPI-549- versus vehicle-treated BALB/c NU/NU mice ($n = 10$). **c**, Mean tumour volume of subcutaneous CT-26 tumour in IPI-549- versus vehicle-treated BALB/c

mice with or without CD8⁺ T cell depletion by anti-CD8 antibody ($n = 10$). **d**, Quantification and representative pictures of CT26 tumour-specific immune responses in PBMCs from CT26 tumour-bearing mice treated with IPI-549 in comparison to vehicle by ELISPOT. PBMCs were collected from tumour-bearing animals after 10 days of vehicle or IPI-549 treatment and restimulated overnight with irradiated CT26 or 4T1 stimulator cells.



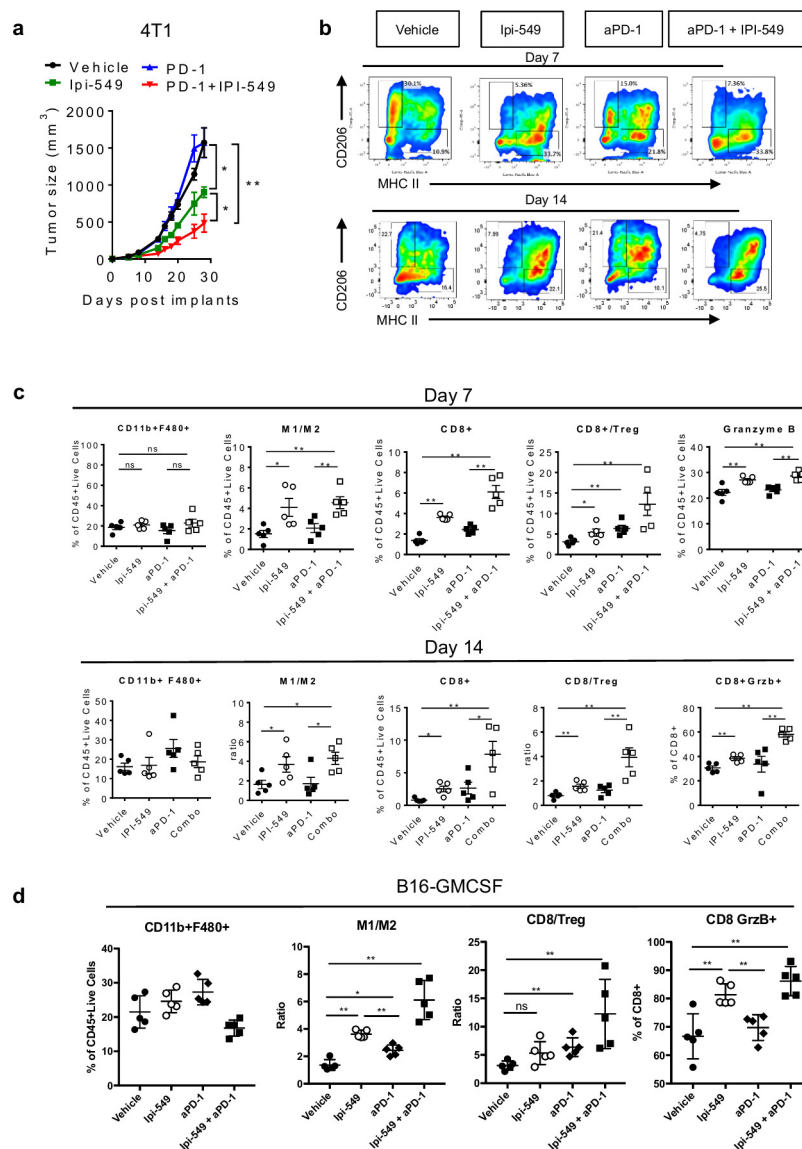
Extended Data Figure 6 | Effect of selective PI3K γ inhibition on the differentiation of T cells in tumours. a, Representative flow cytometry analysis and quantification of CD62L and CD44 expression in CD8⁺ and CD4⁺ T cell infiltrates in tumour, lymph node (LN) and spleen of 4T1

tumour-bearing mice treated with IPI-549 compared to vehicle. Data represent analyses of five mice per group, mean \pm s.e.m. NS, not significant (non-parametric Mann-Whitney *U*-test).



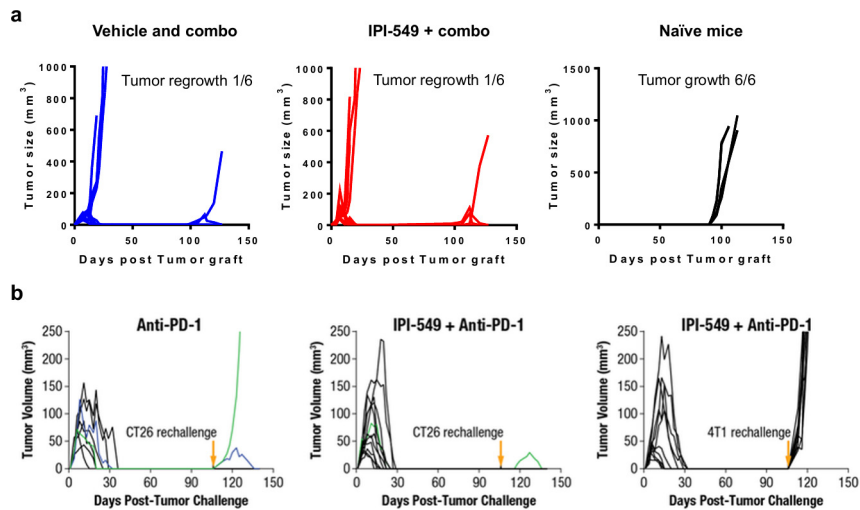
Extended Data Figure 7 | Effect of combination of a selective PI3K γ inhibitor with checkpoint blockade on various tumours. a, Survival to 2,000 mm³ tumour volume of LLC Brei tumour in IPI-549- or vehicle-treated mice in combination with or without anti-CTLA4 (vehicle and

IPI-549 groups, $n = 14$; anti-CTLA4, $n = 13$; IPI-549 and anti-CTLA4 combination, $n = 10$). **b,** Mean tumour volume of CT26 tumour in IPI-549- or vehicle-treated mice in combination with or without anti-PD-L1 ($n = 15$ for all groups except vehicle, $n = 13$).



Extended Data Figure 8 | Effect of combination of a selective PI3K γ inhibitor with checkpoint blockade on TILs. a, Mean tumour volume of subcutaneous 4T1 tumour-bearing mice treated with IPI-549, vehicle or anti-PD-1 in combination with IPI-549 or vehicle ($n = 10$). **b**, Representative flow cytometry analysis of CD206 and MHCII labelling in CD11b⁺ F4/80⁺ cell populations in the different treatment groups of 4T1 tumour-bearing mice. **c**, Quantification of CD11b⁺ F4/80⁺, M1/M2

ratio, CD8⁺/T_{reg} in TILs and granzyme B expression in CD8⁺ T cells from 4T1 tumours in the different treatment groups. **d**, Quantification of CD11b⁺ F4/80⁺, M1/M2 ratio, CD8⁺/T_{reg} in TILs and granzyme B expression in CD8⁺ T cells from B16-GMCSF tumours in the different treatment groups, mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ (non-parametric Mann-Whitney U -test).



Extended Data Figure 9 | Effect of combination of a selective PI3K γ inhibitor with checkpoint blockade on acquisition of anti-tumour memory. **a**, Tumour re-challenge at 100 days (from first tumour implant) following primary tumour complete response in B16-GMCSF tumour-bearing mice treated with vehicle (blue) or IPI-549 (red) in combination with both anti-PD1 and anti-CTLA4. **b**, CT26 tumour-bearing mice with

complete responses in the anti-PD-1 treatment group and the IPI-549 plus anti-PD-1 combination treatment group were re-challenged with CT26 tumour implant. Additional mice with complete responses from the IPI-549 plus anti-PD-1 combination were implanted with 4T1 tumours. There was a low or no tumour take with CT26 re-challenge, while all 4T1 tumours grew, indicating specific anti-tumour memory.

Alternative modes of client binding enable functional plasticity of Hsp70

Alireza Mashaghi^{1*}, Sergey Bezrukavnikov^{1*}, David P. Minde¹, Anne S. Wentink^{2,3}, Roman Kityk², Beate Zachmann-Brand^{2,3}, Matthias P. Mayer², Günter Kramer^{2,3}, Bernd Bukau^{2,3} & Sander J. Tans¹

The Hsp70 system is a central hub of chaperone activity in all domains of life. Hsp70 performs a plethora of tasks, including folding assistance, protection against aggregation, protein trafficking, and enzyme activity regulation^{1–5}, and interacts with non-folded chains, as well as near-native, misfolded, and aggregated proteins^{6–10}. Hsp70 is thought to achieve its many physiological roles by binding peptide segments that extend from these different protein conformers within a groove that can be covered by an ATP-driven helical lid^{11–15}. However, it has been difficult to test directly how Hsp70 interacts with protein substrates in different stages of folding and how it affects their structure. Moreover, recent indications of diverse lid conformations in Hsp70–substrate complexes raise the possibility of additional interaction mechanisms^{15–18}. Addressing these issues is technically challenging, given the conformational dynamics of both chaperone and client, the transient nature of their interaction, and the involvement of co-chaperones and the ATP hydrolysis cycle¹⁹. Here, using optical tweezers, we show that the bacterial Hsp70 homologue (DnaK) binds and stabilizes not only extended peptide segments, but also partially folded and near-native protein structures. The Hsp70 lid and groove act synergistically when stabilizing folded structures: stabilization is abolished when the lid is truncated and less efficient when the groove is mutated. The diversity of binding modes has important consequences: Hsp70 can both stabilize and destabilize folded structures, in a nucleotide-regulated manner; like Hsp90 and GroEL, Hsp70 can affect the late stages of protein folding; and Hsp70 can suppress aggregation by protecting partially folded structures as well as unfolded protein chains. Overall, these findings in the DnaK system indicate an extension of the Hsp70 canonical model that potentially affects a wide range of physiological roles of the Hsp70 system.

To study the DnaK system at the single-molecule level, we first used an engineered substrate that is prone to misfolding *in vitro*: four maltose binding proteins linked head-to-tail (4MBP, Fig. 1a). Mechanical stretching of 4MBP in the absence of chaperone^{20,21} first produced a gradual unfolding of the C-terminal segments of all four MBP proteins (Fig. 1b, N → 4), followed by the distinct unfolding events of the four remaining core structures (Fig. 1b, 4 → 3 → 2 → 1 → U). Relaxation to low forces and a waiting period of 5 s provided an opportunity to refold. However, while subsequent stretching showed a compact structure had formed, these structures typically failed to unfold (termed tight misfold, Fig. 1b) or unfolded in steps larger than one MBP core (termed weak misfold). Isolated 4MBP thus formed small aggregate structures consisting of more than one MBP repeat (Fig. 1c). Next, we performed these experiments in a buffer containing ATP and the DnaK chaperone system (DnaK, DnaJ, GrpE; Fig. 1d and Extended Data Figs 1–4). Unfolding and relaxation now rarely produced tight misfolds, though weak misfolds still formed (Fig. 1e, f). Moreover, the data showed more unfolding steps corresponding to a single MBP core ($P < 0.05$, Fig. 1e,

green arrows), indicating the native-like refolding of core structures. These results show that the DnaK system can promote the correct refolding of a misfolding- and aggregation-prone protein, consistent with previous bulk studies of Hsp70^{2,5,13}. However, they do not reveal how the various components of the DnaK system affect client conformations during the different stages of folding.

To understand the underlying mechanisms we focused on the simpler case of a single MBP to preclude aggregation and omitted the co-chaperones (Fig. 2a). Multiple protocols were followed. First, we unfolded MBP, which occurred in a two-step manner via the core

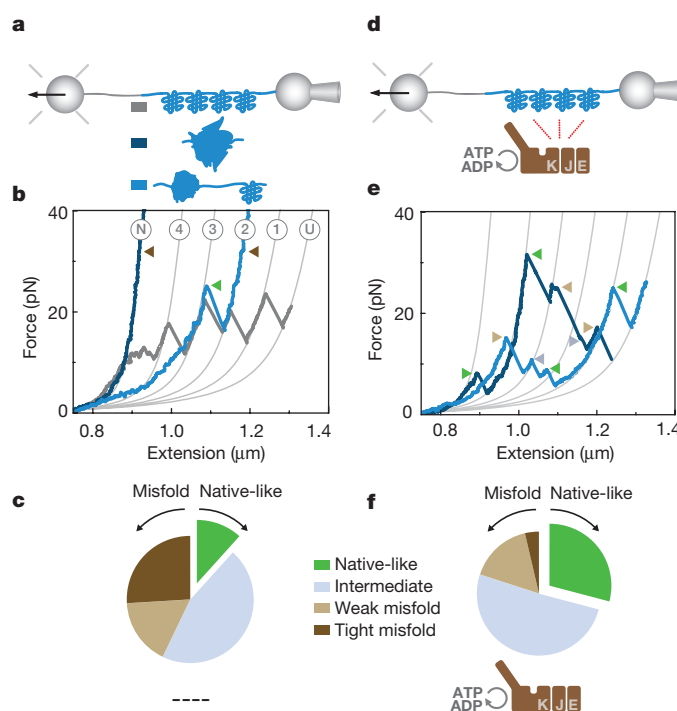


Figure 1 | DnaK chaperone system suppresses aggregation and promotes refolding. **a**, Schematic diagram of experiment. Tandem of MBP monomers (4MBP, blue) is attached to DNA linker (grey) and tethered between beads²⁰. **b**, 4MBP in the absence of chaperones. Thick grey line, first stretching curve showing native unfolding. Dark and light blue lines, subsequent stretching curves showing different misfolded states. Triangles, types of events or states (see panels c and f and main text). Thin grey lines, theoretical compliance for native (N) to unfolded state (U). Numbers of folded core structures are indicated. **c**, Corresponding event fractions ($n = 38$). Intermediate, unfolding events with unfolded length shorter than one MBP core. **d**, Schematic diagram of experiment. **e**, Stretching curves of 4MBP in presence of DnaK system and ATP. **f**, Corresponding event fractions ($n = 55$).

¹FOM institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands. ²Center for Molecular Biology of Heidelberg University (ZMBH), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany. ³German Cancer Research Center (DKFZ), Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany.

*These authors contributed equally to this work.

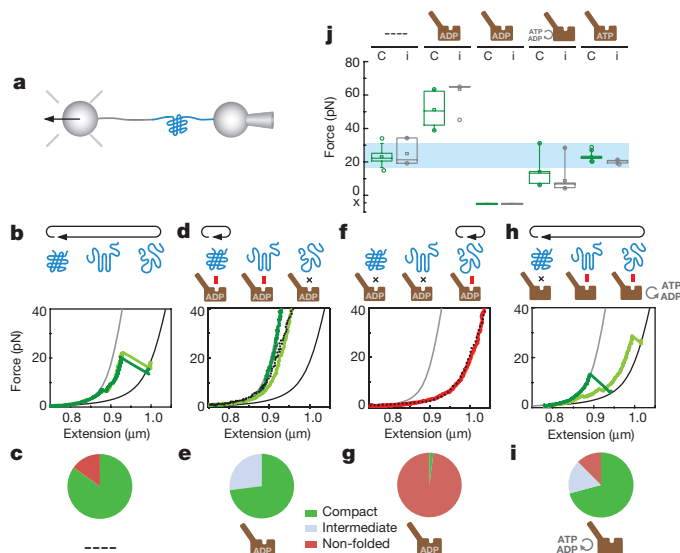


Figure 2 | DnaK binds and stabilizes folded structures. **a**, Schematic diagram of the experiments. **b, c**, Stretching of MBP that was unfolded and relaxed in isolation ($n = 50$). **d, e**, Stretching of MBP that was unfolded and relaxed in isolation and then incubated at low force with $1 \mu\text{M}$ DnaK and ADP loading buffer ($n = 35$). Dotted line, 100 nM DnaK. **f, g**, Stretching of MBP that was incubated in the unfolded state with $1 \mu\text{M}$ DnaK and ADP loading buffer and then relaxed ($n = 15$). Dotted line, 100 nM DnaK. **h, i**, Stretching of MBP that was unfolded and relaxed with DnaK and ATP ($n = 24$). Pie charts indicate states observed after low-force period: compact (native or core-like), intermediate between core and unfolded, and unfolded. Grey and black lines, theoretical compliance for folded and unfolded MBP. **j**, Unfolding forces for core (C) and intermediate (i) states for panels **b, d, f** and **h**. Last column, DnaK trapped in ATP state (Extended Data Fig. 5j, k; $n = 18$). Whiskers, 90% and 10% extreme values; central band, median; rectangle symbol, mean.

state^{20,21} (Fig. 2b), and then performed cycles of relaxation, waiting at 0 pN for 5 s, and subsequent stretching, all in the absence of chaperones. Stretching indicated the chain either had remained unfolded, or had refolded to a compact structure that typically unfolded in the same two-step manner (Fig. 2b). The refolding probability P_r was about 0.85 (Fig. 2c), which increased for longer waiting periods at 0 pN^{20,21}. Second, the cycle was paused at 0 pN, DnaK was added and allowed to interact. The DnaK lid is thought to be predominantly open in the presence of ATP and mostly closed after hydrolysis to ADP. We thus used 1 mM ADP buffer with small amounts ($\sim 5\%$) of ATP (termed ADP loading buffer) in order to stall the ATP cycle after hydrolysis and promote DnaK–MBP binding. Stretching–relaxation cycles were then resumed for the same molecule. Surprisingly, the protein structures now typically failed to unfold over multiple such cycles until the tether broke. The structures were stable at high forces (maximally sustained force $F_u > 40 \text{ pN}$) well above the native unfolding force ($22 \pm 5 \text{ pN}$), and sometimes even exceeded the effective maximum for our assay (65 pN)²² (Fig. 2d, j).

Notably, this scenario is unlike 4MBP aggregation (Fig. 1b), as aggregation partners are not available here. Instead, the data suggest that in ADP loading buffer, DnaK bound to folded structures and stabilized them against forced unfolding. When MBP molecules were stretched for the first time in the presence of DnaK and ADP loading buffer, they unfolded in similar manner as isolated MBP without DnaK ($F_u = 23 \pm 4 \text{ pN}$), which indicates that DnaK does not interact with fully native MBP, or only weakly. Rather, the data suggest that MBP is in a near-native state when stabilized by DnaK. Specifically, the MBP states stabilized after DnaK addition (Fig. 2d) had folded in the absence of chaperones, which yields native-like states (Fig. 2b). The stabilized structures were typically compact, much like the native or core MBP states, but could also display lengths in-between those of the core and unfolded states (Fig. 2d, e). Native MBP could also be stabilized after a minimal

perturbation, by unfolding the C-terminal segment in a first pull and then relaxing the force to preserve the native core state (Extended Data Fig. 5a, b). Overall, these data indicate that DnaK can stabilize partially folded and near-native MBP states against forced unfolding.

Previous work had suggested, however, that DnaK rather stabilizes the unfolded state of substrates (Extended Data Figs 2–4)²³. With the aim of promoting this binding mode, we modified the protocol. After a few unfolding–refolding cycles, unfolded MBP was kept at a moderate force to prevent refolding, and incubated for 1 to 2 minutes with DnaK and ADP loading buffer. In subsequent relaxation–stretching cycles, MBP now remained unfolded ($P_r \approx 0.05$; Fig. 2f, g), indicating that DnaK had bound to and stabilized the unfolded chain. Next, we reduced the DnaK incubation time from minutes to seconds by first unfolding MBP in the presence of DnaK and ADP loading buffer, and immediately subjecting it to relaxation and stretching (Extended Data Fig. 5c, d). Unfolded chains now did refold in subsequent cycles. Moreover, the refolded structures typically unfolded at increasingly high forces, until they were locked into a compact state that could not be unfolded (Extended Data Fig. 5d). Consistently, such a compact and locked state was achieved in earlier cycles for longer waiting times at 0 pN (Extended Data Fig. 5e, f). These data suggest a kinetic partitioning: unfolded chains can be bound by DnaK²⁴, which blocks their refolding, unless the chains fold first (folding time is about 1 s for MBP)²⁰ and are then stabilized by DnaK in a folded state.

The observed stabilization by DnaK also occurred during active ATP hydrolysis. In the presence of DnaK and 1 mM ATP, we found that refolded structures frequently unfolded via partially folded structures smaller than the core state (44% of the traces, lifetime of partial structure $\tau = 1.1 \pm 0.6 \text{ s}$, Fig. 2h, i). For example, the light green curve in Fig. 2h indicates a partially folded structure beyond 5 pN that was stable for $\tau \approx 4 \text{ s}$, and then unfolded fully at 30 pN. This observation suggested the folded structure was stabilized by DnaK, and then unfolded when DnaK was released as it progressed through the ATP cycle. Consistently, the occurrence of transiently stable partial folds was higher than without chaperones ($P < 0.01$, 8% of the traces, $\tau \approx 0.5 \text{ s}$), but lower than with DnaK and ADP loading buffer ($P < 0.01$, all traces, $\tau \leq 150 \text{ s}$). The stabilization of extended and folded states by DnaK was also observed at lower DnaK concentrations (Fig. 2d, f and Extended Data Fig. 5g), and in presence of HPLC purified (ATP free) ADP (Extended Data Fig. 5h, i). Consistently, stabilization was not observed for a DnaK mutant (T199A) that is trapped in the open ATP state (Fig. 2j and Extended Data Fig. 5j, k). Overall, these results indicate that stabilization of partially folded structures by DnaK is modulated by its ATP hydrolysis cycle.

We next attempted to identify which structural elements of the Hsp70 substrate-binding domain are conducive to stabilizing folded MBP conformations. The helical lid may have a role, as implied by the observation that ATP affects both binding to folded structures (Fig. 2d, h) and lid conformation¹⁶, but this evidence is indirect. To address this issue more directly, we investigated DnaK with a truncated lid (DnaK(2–538), Extended Data Fig. 6)²⁴. In the presence of this variant and ADP loading buffer, refolded MBP molecules did not show stabilization but rather native-like core unfolding, whether the waiting time at 0 pN was 5 s ($F_u = 22 \pm 5 \text{ pN}$, Fig. 3a, e) or minutes ($F_u = 24 \pm 7 \text{ pN}$, Extended Data Fig. 5l). By contrast, stabilization was observed after minute-long exposure at 0 pN to a DnaK variant that has a mutated binding groove but retains the lid (DnaK(V436F), Fig. 3c). Unfolding then occurred at forces up to 65 pN (Fig. 3c, e), and often (52% of the traces) via partial folds that are only scarcely observed without chaperone²². The unfolding forces were lower than for wild-type DnaK however ($P < 0.01$, Fig. 3e). This groove mutation is reported to reduce peptide affinity by more than tenfold²⁴. Consistent with such a peptide-binding defect, we found that after minute-long exposure to groove-mutated DnaK in the extended state, MBP chains formed folded structures when relaxed to 0 pN for 5 s ($n = 13$), while refolding was efficiently blocked in the wild-type DnaK exposure experiments

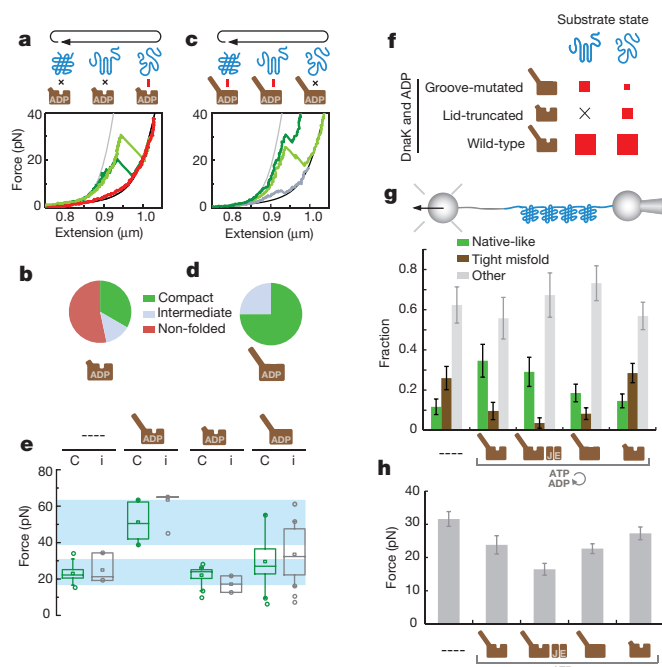


Figure 3 | DnaK lid is central to stabilizing folded structures and suppressing aggregation. **a, b**, Stretching of MBP that was unfolded and relaxed with lid-truncated DnaK(2–538) in ADP loading buffer ($n = 15$). Waiting time at low force is 5 s. **c, d**, Stretching of MBP that was unfolded and relaxed with groove-mutated DnaK(V436F) in ADP loading buffer ($n = 21$). Waiting time at low force is of the order of minutes; pie charts indicate subsequently observed states: compact (native or core-like), intermediate between core and unfolded, and unfolded. **e**, Unfolding forces for core (C) and intermediate (i) states. Whiskers, 90% and 10% extreme values; central band, median; rectangle symbol, mean. **f**, Red squares qualitatively indicate observed stabilization of folded and extended states by DnaK variants. **g**, 4MBP refolding statistics. Indicated is type of event as fraction of all events. Error bars, ± 1 s.d. **h**, Mean unfolding force of non-native structures formed during 4MBP refolding. Error bars, ± 1 s.e.m.

(Fig. 2f, g). Lid-truncated DnaK on the other hand, which retains the wild-type groove, was found to suppress refolding in regular stretch-relax cycles ($P < 0.01$, Fig. 3a, b). Thus, the stabilization of folded structures by DnaK is abolished by truncating the lid and weakened by mutating the groove.

We surmised that binding partially folded structures could help to suppress aggregation^{8,9,25} (Fig. 1e, f). If so, groove-mutated DnaK should retain some of this capability, as it binds folded structures (Fig. 3c). We found that the presence of groove-mutated DnaK and 1 mM ATP indeed decreased tight aggregation of 4MBP ($P < 0.01$, Fig. 3g and Extended Data Fig. 7). The aggregated structures that did form, unfolded at lower forces ($P < 0.01$, Fig. 3h). These findings are consistent with models where DnaK can destabilize misfolded structures⁴. The results may appear paradoxical: how can DnaK both result in more- (Fig. 2d) and less-stable (Fig. 3h) structures? The ATP-driven DnaK cycle offers a possible rationale. First, DnaK stabilizes in the ADP state, and hence only transiently. Moreover, by binding less-stable structures early after relaxation, DnaK can limit more-stable structures that would form later (Fig. 4a). Indeed, aggregates may have different structure and be less stable when formed with DnaK present. DnaK may also discriminate between native and non-native structures, and directly destabilize the latter⁴. We find DnaK selectively stabilizes partially and not fully folded MBP, though DnaK selectivity may go further. Overall, the data indicate that aggregation can be suppressed by the transient stabilization of (partially) folded structures.

Second, other components of the DnaK system may also contribute to limiting aggregation. With the full DnaK system present, tight aggregation was less frequent than for DnaK and ATP, though the

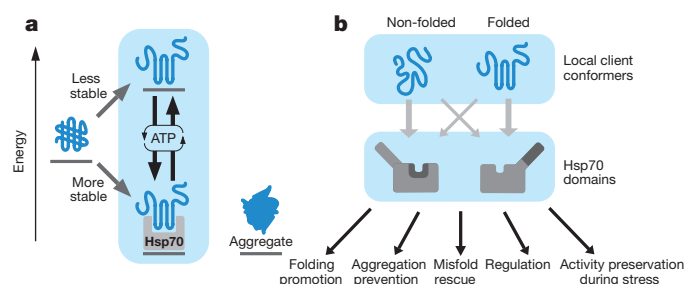


Figure 4 | Model of Hsp70 chaperone action. **a**, Effect on substrate stability. Hsp70 can transiently stabilize partially folded structures that are less stable than complete folds upon Hsp70 release. Hsp70 binding can suppress the formation of permanently stable aggregates. ATP and co-chaperones provide stability control. For instance, high ADP and low ATP levels or low nucleotide exchange rates with heat-denatured GrpE³¹, or Hsp in mammalian cells³⁰, can prolong Hsp70-mediated stabilization. Hsp70 could bind and (de)stabilize certain structural motifs. **b**, Hsp70 binding modes and roles of domains. This binding diversity and resulting ability to stabilize and destabilize folded structures could affect a wide range of Hsp70 physiological roles.

difference was not large ($P < 0.01$, Fig. 3g). The co-chaperone GrpE did not display any substrate interactions by itself (Extended Data Fig. 8e). By contrast, DnaJ alone completely abolished aggregation (Extended Data Fig. 8a, b), but blocked refolding as well (Extended Data Fig. 8c, d). Within the full DnaK chaperone system, the holdase activity of DnaJ did not dominate, as efficient refolding was observed (see Fig. 1e, f). These findings thus support previous observations that with ATP and GrpE, DnaK can promote DnaJ release from substrate²⁶.

Finally, the observations led us to speculate that DnaK could preserve protein structure and function under heat stress. As DnaK unfolds at lower temperatures than MBP, the latter is unsuitable. We instead investigated a monomeric variant of the replication initiation protein RepE (RepE54, Extended Data Fig. 4d), a well-established DnaK substrate²⁷, using tryptophan fluorescence. We found that with (tryptophan-free) DnaK and ADP, the apparent thermal stability of RepE54 increased by about 5 °C (Extended Data Fig. 9). This upshift was not observed for DnaK and ATP or for lid-truncated DnaK with ADP, consistent with central roles for the ADP state and the lid in the stabilization of folded states. Groove-mutated DnaK with ADP also did not show an upshift, suggesting that the groove was more important here than in the single-molecule MBP experiments. We further observed that DnaK suppressed the temperature-induced decrease of luciferase activity in the presence of ADP (Extended Data Fig. 4c), whereas the ADP buffer alone did not (Extended Data Fig. 10). Together, these data suggest that DnaK can stabilize protein structure and activity during heat stress when ATP abundance is limited⁵.

From these experiments on DnaK, a picture emerges of an Hsp70 functional repertoire that is broader than previously assumed (Fig. 4b): 1) Hsp70 not only binds extended peptides^{4,15,16,25}, but also directly binds (partially) folded structures. The Hsp70 lid is central and could adopt reported open conformations^{15–17,19} to accommodate and bind folded structures, though indirect action cannot be excluded, while the groove also conveys stability. 2) In the canonical model, protein chains fold autonomously after Hsp70 release, and can subsequently be transferred to other chaperones that act in later stages, such as Hsp90 and GroEL^{11,15}. Our findings suggest Hsp70 can also guide and organize late stages of folding¹⁸, for instance by limiting inter-domain contacts²², and may not fully release protein substrates until near native. 3) Hsp70 can protect partially folded structures against aggregation in addition to extended peptide segments²⁸. 4) Hsp70 is known to destabilize (mis)folded protein structures⁴. Hsp70 could achieve this role by interacting directly with misfolded and aggregated structures. Furthermore, our data show that Hsp70 can also do the opposite, and stabilize partially folded protein structures. These contrasting effects may be controlled by ADP abundance and co-chaperones, which can

for instance modulate the lifetime of the ADP state. 5) These findings suggest that Hsp70 can preserve enzymatic functions during stress at low energy costs, in line with the reduced energy availability in episodes of stress.

These Hsp70 functions could have important consequences for our understanding of normal cellular physiology as well as pathophysiology. Hsp70 may transfer clients to other chaperones^{8,9} without requiring (local) unfolding, and contribute to an evolutionary capacitor role by stabilizing substrate proteins²⁹. Co-factors such as HIP³⁰ that promote the ADP state could have a role in regulating client stabilization. Finally, Hsp70 could control signalling pathways⁶ by interacting directly with the folded parts of transcription factors, kinases, and receptors, in addition to their unfolded segments^{1,3}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 July 2015; accepted 4 October 2016.

Published online 26 October 2016.

- Rodriguez, F. *et al.* Molecular basis for regulation of the heat shock transcription factor sigma32 by the DnaK and DnaJ chaperones. *Mol. Cell* **32**, 347–358 (2008).
- Skowrya, D., Georgopoulos, C. & Zylicz, M. The *E. coli* dnaK gene product, the hsp70 homolog, can reactivate heat-inactivated RNA polymerase in an ATP hydrolysis-dependent manner. *Cell* **62**, 939–944 (1990).
- Kirschke, E., Goswami, D., Southworth, D., Griffin, P. R. & Agard, D. A. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell* **157**, 1685–1697 (2014).
- Sharma, S. K., De los Rios, P., Christen, P., Lustig, A. & Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat. Chem. Biol.* **6**, 914–920 (2010).
- Schröder, H., Langer, T., Hartl, F. U. & Bukau, B. DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *EMBO J.* **12**, 4137–4144 (1993).
- Mayer, M. P. & Bukau, B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **62**, 670–684 (2005).
- Frydman, J., Nimmegern, E., Ohtsuka, K. & Hartl, F. U. Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370**, 111–117 (1994).
- Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
- Tyedmers, J., Mogk, A. & Bukau, B. Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* **11**, 777–788 (2010).
- Warrick, J. M. *et al.* Suppression of polyglutamine-mediated neurodegeneration in *Drosophila* by the molecular chaperone HSP70. *Nat. Genet.* **23**, 425–428 (1999).
- Zhu, X. *et al.* Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606–1614 (1996).
- Rüdiger, S., Germeroth, L., Schneider-Mergener, J. & Bukau, B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.* **16**, 1501–1507 (1997).
- Szabo, A. *et al.* The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK, DnaJ, and GrpE. *Proc. Natl Acad. Sci. USA* **91**, 10345–10349 (1994).
- Qi, R. *et al.* Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP. *Nat. Struct. Mol. Biol.* **20**, 900–907 (2013).
- Kityk, R., Kopp, J., Sinning, I. & Mayer, M. P. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Mol. Cell* **48**, 863–874 (2012).
- Schlecht, R., Erbse, A. H., Bukau, B. & Mayer, M. P. Mechanics of Hsp70 chaperones enables differential interaction with client proteins. *Nat. Struct. Mol. Biol.* **18**, 345–351 (2011).
- Smock, R. G., Blackburn, M. E. & Gierasch, L. M. Conserved, disordered C terminus of DnaK enhances cellular survival upon stress and DnaK *in vitro* chaperone activity. *J. Biol. Chem.* **286**, 31821–31829 (2011).
- Buchberger, A., Schröder, H., Hestekamp, T., Schönfeld, H. J. & Bukau, B. Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. *J. Mol. Biol.* **261**, 328–333 (1996).
- Marcinowski, M. *et al.* Substrate discrimination of the chaperone BiP by autonomous and cochaperone-regulated conformational transitions. *Nat. Struct. Mol. Biol.* **18**, 150–158 (2011).
- Bechtluft, P. *et al.* Direct observation of chaperone-induced changes in a protein folding pathway. *Science* **318**, 1458–1461 (2007).
- Mashaghi, A. *et al.* Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **500**, 98–101 (2013).
- Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1997).
- Sikor, M., Mapa, K., von Voithenberg, L. V., Mokranjac, D. & Lamb, D. C. Real-time observation of the conformational dynamics of mitochondrial Hsp70 by spFRET. *EMBO J.* **32**, 1639–1649 (2013).
- Mayer, M. P. *et al.* Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol.* **7**, 586–593 (2000).
- Kellner, R. *et al.* Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein. *Proc. Natl Acad. Sci. USA* **111**, 13355–13360 (2014).
- Gamer, J. *et al.* A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor sigma32. *EMBO J.* **15**, 607–617 (1996).
- Ishiai, M., Wada, C., Kawasaki, Y. & Yura, T. Replication initiator protein RepE of mini-F plasmid: functional differentiation between monomers (initiator) and dimers (autogenous repressor). *Proc. Natl Acad. Sci. USA* **91**, 3839–3843 (1994).
- Nunes, J. M., Mayer-Hartl, M., Hartl, F. U. & Müller, D. J. Action of the Hsp70 chaperone system observed with single proteins. *Nat. Commun.* **6**, 6307 (2015).
- Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- Li, Z., Hartl, F. U. & Bracher, A. Structure and function of Hip, an attenuator of the Hsp70 chaperone cycle. *Nat. Struct. Mol. Biol.* **20**, 929–935 (2013).
- Grimshaw, J. P., Jelesarov, I., Schönfeld, H. J. & Christen, P. Reversible thermal transition in GrpE, the nucleotide exchange factor of the DnaK heat-shock system. *J. Biol. Chem.* **276**, 6098–6104 (2001).

Acknowledgements Work in the group of S.T. is supported by the Foundation for Fundamental Research on Matter (FOM) and the Netherlands Organization for Scientific Research (NWO). Work in the laboratory of B.B. was supported by research grants from the Deutsche Forschungsgemeinschaft (SFB638 and FOR1805) to G.K. and B.B. The work in the laboratory of M.P.M. was funded by the Deutsche Forschungsgemeinschaft (MA 1278/4-1). We thank M. M. Naqvi for performing control experiments, M. Avellaneda for help with preparing protein structure illustrations, and T. Shimizu, E. Garnett and F. Huber for critical reading of the manuscript.

Author Contributions A.M., S.B., D.M., M.P.M., G.K., B.B. and S.T. conceived and designed the research; B.Z.B. and D.M. designed and purified the MBP protein constructs; A.M. and S.B. performed the optical tweezers experiments; B.Z.B., G.K., A.W. and R.K. purified the DnaK system chaperones and DnaK mutants; A.W. performed the bulk RepE54 assays; R.K. performed the bulk luciferase assays; A.M., S.B., A.W. and S.T. analysed the data; and A.M., S.B., D.M., M.P.M., G.K., B.B. and S.T. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.T. (tans@amolf.nl).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Expression and purification of MBP and 4MBP. N-terminally biotinylated MBP and 4MBP C-terminally fused with 4 Myc tag sequences were produced in *Escherichia coli* as hybrid proteins consisting of an N-terminal Ulp1-cleavable N-terminal His₁₀-SUMO sequence followed by an AviTag sequence (Avidity, LCC, Aurora, Colorado, USA), facilitating *in vivo* biotinylation and four consecutive C-terminal Myc-tag sequences. Proteins were purified from *E. coli* BL21(DE3) cells harbouring pBirAcm encoding the biotin ligase (Avidity, LCC, Aurora, Colorado, USA). For overexpression overnight cultures were diluted 1:100 in fresh LB medium supplemented with 20 mg/l biotin, 20 mg/l kanamycin, 10 mg/l chloramphenicol, 0.2% glucose and incubated under vigorous shaking at 30°C. Expression was induced at OD₆₀₀ = 0.6 by addition of 1 mM IPTG for 3 h. Cells were chilled, harvested by centrifugation, flash-frozen in liquid nitrogen and stored at -70°C. Cell pellets were resuspended in ice-cold buffer A (20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 1% Triton X-100, 1 mM PMSF) and lysed using a French pressure cell. The lysate was cleared from cell debris by centrifugation at 35,000g for 30 min and incubated with Ni-IDA matrix (Protino; Macherey-Nagel, Düren, Germany) for 30 min at 4°C. The matrix was washed extensively with buffer A and bound hybrid proteins were eluted in buffer A containing 250 mM imidazole. The eluate was supplemented with His6-Ulp1 protease and dialysed overnight at 4°C in buffer D (20 mM Tris-HCl pH 7.5, 0.2 M NaCl). Following dialysis coupled with Ulp1 digestion, His6-Ulp1 protease and the His₁₀-SUMO fragment were removed by incubation with Ni-IDA matrix. The unbound fraction containing cleaved MBP or 4MBP was then loaded on Amylose resin (New England Biolabs) previously equilibrated in buffer D, washed with cold buffer D and bound proteins were eluted in buffer D supplemented with 20 mM maltose. Elution fractions were dialysed three times for 2 h at 4°C in 100-fold excess volume of buffer S (20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 1 mM EDTA). 4MBP purifications in addition were subjected to size-exclusion chromatography using a HiLoad 16/600 Superdex prep grade column. Purified proteins were concentrated using Vivaspin centrifugal concentrators, aliquoted, flash frozen in liquid nitrogen and stored at -70°C.

Biotinylated proteins were also produced as follows. Purification of MBP and 4MBP containing one N-terminal cysteine was performed as described²⁰. For biotinylation the purified proteins were dialysed in buffer B (20 mM Tris-HCl pH 7.5, 0.2 M NaCl, 1 mM EDTA) and incubated on ice for 15 min in buffer B containing 5 mM TCEP to reduce disulphide bonds. A tenfold molar excess of maleimide-PEG11-biotin (Thermo Scientific) dissolved in DMSO was added and biotinylation was performed for 2 h at 25°C. Proteins were concentrated and subjected to size-exclusion chromatography using a HiLoad 16/600 Superdex prep grade column. Purified fractions were analysed by SDS-PAGE, pooled, aliquoted, flash frozen in liquid nitrogen and stored at -70°C.

Wild-type and mutant chaperone expression and purification. All proteins were produced in *E. coli* from expression plasmids. Wild-type DnaK, DnaK(2-538) (lid-truncated), and DnaK-V436F (groove-mutated) were produced in a Δ dnaK strain and purified by ammonium sulphate precipitation, anion exchange chromatography (DEAE-sepharose), ATP-agarose affinity chromatography, gel-filtration and strong anion exchange chromatography (Resource Q) as described earlier³². DnaJ was purified by chromatography on cation exchange and hydroxyapatite material. GrpE was expressed in a Δ dnaK strain and purified by chromatography on DEAE-Sepharose, hydroxyapatite and Superdex 200 as described³³. Tryptophan-free DnaK mutants (W102F) were expressed with an N-terminal His₁₀-SUMO sequence and purified by Ni-affinity chromatography, Ulp1 digestion and a second Ni-affinity chromatography step.

Luciferase activity protection assay. Luciferase (80 nM in 40 mM HEPES/KOH pH 7.5, 50 mM potassium acetate, 5 mM magnesium acetate, 2 mM DTT) was incubated in the absence or presence of ATP (2 mM), ADP (3.2 μ M), wild-type or mutant DnaK proteins (800 nM), DnaJ (160 nM) and GrpE (400 nM) at 37°C. Luciferase activity was determined by diluting 1 μ l aliquots into 124 μ l assay buffer (100 mM potassium phosphate (pH 7.8), 25 mM glycylglycine (pH 7.4), 100 mM potassium acetate, 15 mM magnesium acetate, 5 mM ATP), mixing with 125 μ l luciferin (80 μ M in assay buffer) and monitoring bioluminescence for 5 s without delay using a Biolumat (Berthold).

RepE54 thermal unfolding assay. RepE54 (2 μ M in 50 mM HEPES/KOH pH 7.5, 50 mM KCl, 5 mM MgCl₂, 2 mM DTT) was incubated at 25°C in the absence or presence of ATP (2 mM), ADP (2 mM), ADP loading buffer conditions (2 mM ADP + 5% ATP), tryptophan-free DnaK wild-type or mutant proteins (2 μ M) for 30 min. The unfolding transition of RepE54 was then monitored by fluorescence emission at 329 nm (excitation at 295 nm) as a function of temperature

(temperature gradient from 25–65°C at a rate of 90°C/hour) on a Jasco FP-6500 spectrofluorometer.

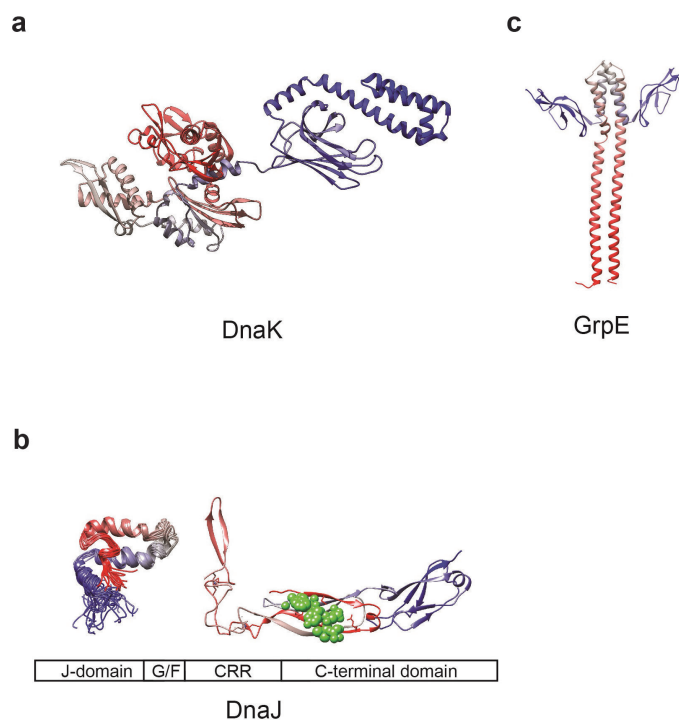
Optical tweezers assay. Anti-digoxigenin (anti-Dig) coated beads (DIGP-20-2, diameter 2.1 μ m) were purchased from SpheroTech and stored at 4°C until use. 2,553-base-pair DNA handles were prepared and functionalized with digoxigenin (Dig) and biotin at the 5'-ends of both strands as described previously²¹. DNA-coated microspheres were made by mixing ~70 ng of biotin-dsDNA-Dig and 4 μ l Dig-coated beads in 10 ml HMK buffer (50 mM HEPES, pH 7.5, 5 mM MgCl₂, 100 mM KCl). After a 30 min incubation on a rotary mixer (4°C), 1 μ l of neutravidin solution (1% w/v) was added to the mixture and incubated at 4°C for 10 more minutes. Next, the unbound neutravidin was washed away by centrifugation, and the beads were dissolved in 400 μ l HMK buffer for use in optical tweezers experiments. Carboxylated polystyrene beads (CP-20-10, diameter 2.1 μ m, SpheroTech) were covalently attached to anti-Myc antibody (Roche Diagnostics) via carbodiimide reaction (Poly-Link Protein Coupling Kit, Polysciences Inc.). Briefly, the beads were washed and then mixed with freshly prepared 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide and the antibody, and the mixture was incubated for 3 h. MBP-coated beads were made by mixing 5 μ l of 50 μ M solution of either 1MBP or 4MBP and 2 μ l Myc-coated beads in 10 μ l HMK buffer and incubating the mixture for 30 min on a rotary mixer at 4°C. Subsequently, the beads were dissolved in 400 μ l HMK buffer for use in optical tweezers experiments.

Optical tweezers assays, buffer contents: Fig. 1b, c, no chaperones; Fig. 1e, f, 100 nM DnaK, 100 nM DnaJ, 50 nM GrpE, 1 mM ATP; Fig. 2b, c, no chaperones; Fig. 2d–g, 1 μ M DnaK, 1 mM ADP loading buffer; Fig. 2h, i, 100 nM DnaK, 1 mM ATP; Fig. 3a, b, 1 μ M DnaK lid-truncated, 1 mM ADP loading buffer; Fig. 3c, d, 1 μ M DnaK groove-mutated, 1 mM ADP loading buffer; Extended Data Fig. 5b, 100 nM DnaK, 1 mM ADP loading buffer; Extended Data Fig. 5c–f, 1 μ M DnaK, 1 mM ADP loading buffer; Extended Data Fig. 5g, 100 nM DnaK, 1 mM ADP loading buffer; Extended Data Fig. 5h, i, 1 μ M DnaK, 100 μ M ADP purified by HPLC; Extended Data Fig. 5j, k, 100 nM DnaK T199A, 1 mM ATP; Extended Data Fig. 5l, 1 μ M lid-truncated DnaK and 1 mM ADP loading buffer; Extended Data Fig. 7b, 100 nM DnaK, 1 mM ATP; Extended Data Fig. 7c, 100 nM DnaK lid-truncated, 1 mM ATP; Extended Data Fig. 7d, 100 nM DnaK groove-mutated, 1 mM ATP; Extended Data Fig. 8b, 1 μ M DnaJ; Extended Data Fig. 8c, 100 nM DnaJ; Extended Data Fig. 8d, 1 μ M GrpE. Number of events (high force): $n = 38$ (Fig. 1b, c), 55 (Fig. 1e, f), 52 (Extended Data Fig. 7b), 131 (Extended Data Fig. 7c), 96 (Extended Data Fig. 7d). Number of events (low force): $n = 50$ (Fig. 2b, c), 35 (Fig. 2d, e), 15 (Fig. 2f, g), 24 (Fig. 2h, i), 15 (Fig. 3a, b), 21 (Fig. 3c, d), 6 (Extended Data Fig. 5b), 10 (Extended Data Fig. 5c, d), 11 (Extended Data Fig. 5e, f), 17 (Extended Data Fig. 5g, h), 15 (Extended Data Fig. 5i, j), 4 (Extended Data Fig. 8b), 9 (Extended Data Fig. 8c), 10 (Extended Data Fig. 8d). ADP loading buffer was prepared by dissolving ADP sodium salt (A2754, Sigma Aldrich) in HMK buffer. The ATP fraction in this buffer was ~5%, as measured by HPLC.

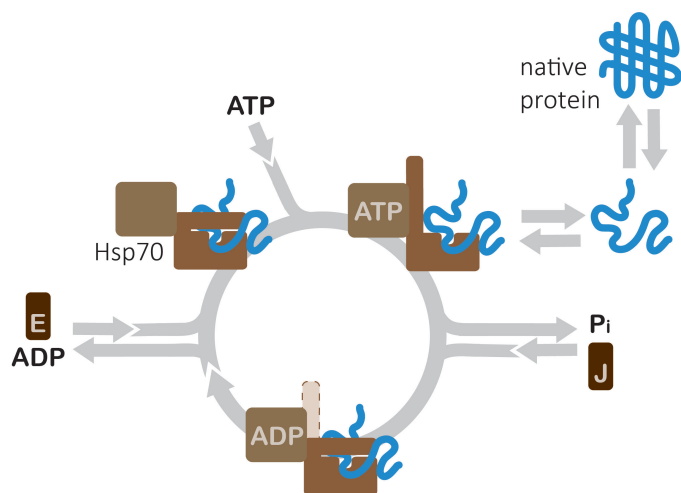
Stretching experiments were performed at room temperature using a custom made optical tweezers setup¹⁸. An MBP-coated bead was trapped in the optical trap and then transferred to a micropipette tip, and, subsequently, a DNA-coated bead was trapped. Next, two beads were brought in close contact, allowing a tether between the beads to form. When the micropipette is moved away from the trap, the tether experiences a stretching force, leading to its extension and deflection of the trapped bead from the trap centre. Specific tethering to natively folded MBP molecules was detected by the typical two-step unfolding upon stretching. Tethers that broke at low force or showed no unfolding, for instance because of non-specific binding of the neutravidin at the end of the DNA tether to the bead surface, were excluded. The deflection is monitored by collecting the transmitted laser light and directing it to quadrant photodiode where it is recorded at 50 Hz. The data are filtered with 5th order Butterworth filter at 20 Hz before saving. Trap stiffness and sensitivity were measured to be 169 ± 24 pN/ μ m and 2.74 ± 0.24 V/ μ m correspondingly. A nanopositioning piezo stage is used to move the flow chamber and micropipette at a speed of 50 nm/s, which corresponds to a pulling rate on the tethered MBP construct of ~5 pN/s at unfolding. Differences between pie charts were analysed using one-tailed two-proportion z-test. Statistical analysis of sustained forces in different conditions was performed using the Kruskal–Wallis (one-way ANOVA on ranks) test, given that the forces are not distributed normally and have non-equal variances for different groups. Test results such as *P* values are reported in the main text. In box plots, whiskers indicate 90% and 10% extreme values, within the box the band indicates the median, rectangle indicates the mean. In Fig. 3g, error bars are ± 1 s.d. In Fig. 3h, error bars are ± 1 s.e.m.

Data availability. The data that support the findings of this study are available from the authors on reasonable request.

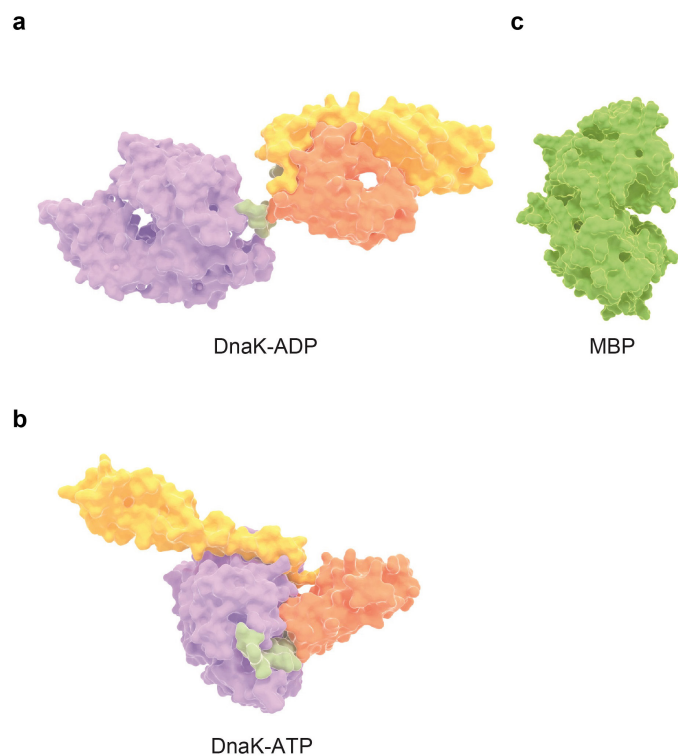
32. Buchberger, A., Valencia, A., McMacken, R., Sander, C. & Bukau, B. The chaperone function of DnaK requires the coupling of ATPase activity with substrate binding through residue E171. *EMBO J.* **13**, 1687–1695 (1994).
33. Packschies, L. *et al.* GrpE accelerates nucleotide exchange of the molecular chaperone DnaK with an associative displacement mechanism. *Biochemistry* **36**, 3417–3422 (1997).
34. Pellecchia, M., Szyperski, T., Wall, D., Georgopoulos, C. & Wüthrich, K. NMR structure of the J-domain and the Gly/Phe-rich region of the Escherichia coli DnaJ chaperone. *J. Mol. Biol.* **260**, 236–250 (1996).
35. Li, J., Qian, X. & Sha, B. The crystal structure of the yeast Hsp40 Ydj1 complexed with its peptide substrate. *Structure* **11**, 1475–1483 (2003).
36. Harrison, C. J., Hayer-Hartl, M., Di Liberto, M., Hartl, F. & Kuriyan, J. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science* **276**, 431–435 (1997).
37. Mayer, M. P. Gymnastics of molecular chaperones. *Mol. Cell* **39**, 321–331 (2010).
38. Bertelsen, E. B., Chang, L., Gestwicki, J. E. & Zuiderweg, E. R. Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc. Natl Acad. Sci. USA* **106**, 8471–8476 (2009).



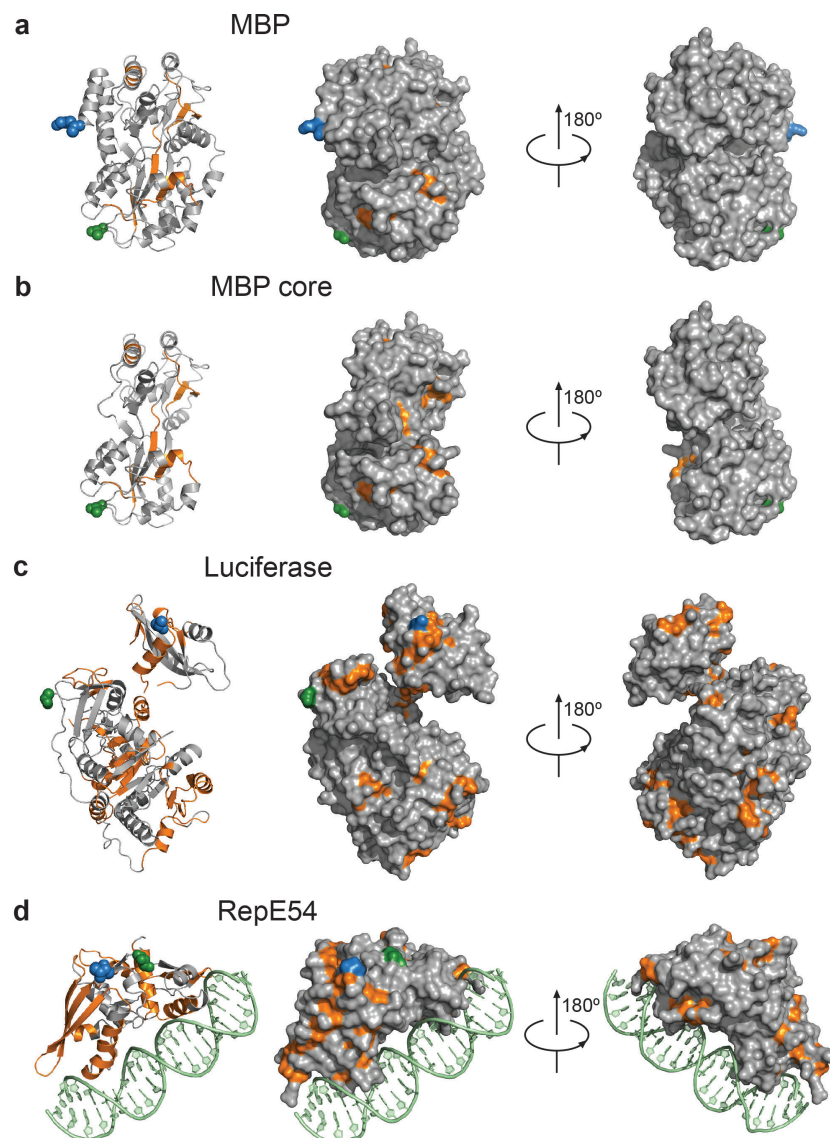
Extended Data Figure 1 | Domain architectures of DnaK, DnaJ and GrpE. **a**, Structure of *E. coli* DnaK bound to peptide substrate and ADP nucleotide (PDB ID: 2KHO) in N-to-C-terminal colour gradient from red (N terminus) through dark grey to blue (C terminus)³³. **b**, DnaJ domain structure containing J domain, glycine/phenylalanine-rich domain (G/F), cysteine-rich region/zinc-finger (CRR) and C-terminal domain; partial NMR structures of J domain (PDB ID: 1XBL) and crystal structure comprising cysteine-rich domain (CRR) and peptide-binding domain of the yeast homologue with co-crystallized peptide substrate highlighted in green spheres (PDB ID: 1NLT)^{34,35}. **c**, GrpE structure (PDB ID: 1DKG)³⁶.



Extended Data Figure 2 | Canonical nucleotide-driven cycle of Hsp70–client interactions. Hsp70 is known to interact with a large range of protein conformations from unfolded nascent chains to near-native proteins to aggregates and misfolded states. Non-native protein conformers are captured via short extended peptide segments by the substrate-binding groove of Hsp70. Hsp40 (DnaJ) then accelerates the ATP hydrolysis reaction together with the bound substrate. Hsp70–ADP locks the substrate under the closed helical lid of the substrate-binding domain in the ‘high-affinity’ state. The lid could also adopt the open conformation according to recent indications^{16–18,20}. Nucleotide exchange factor (GrpE) subsequently accelerates the release of ADP and either spontaneous fluctuations of the lid or new ATP molecules facilitate the release of the substrate to regenerate Hsp70–ATP for another round of the cycle³⁷.

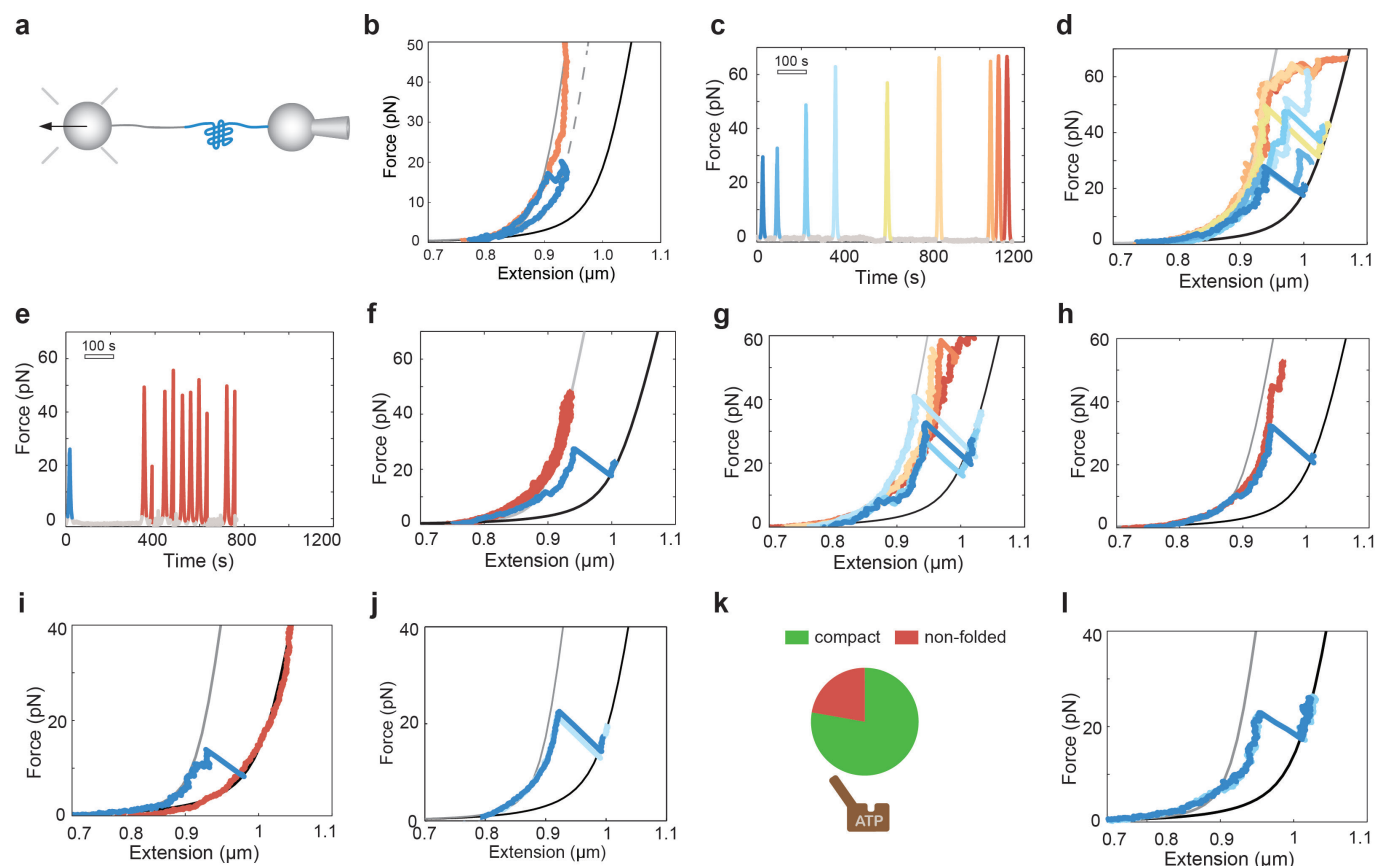


Extended Data Figure 3 | Protein structures. **a**, DnaK in ADP state (PDB ID: 2KHO). Domain boundaries indicated in different colours: nucleotide-binding domain (purple, 1–383), hydrophobic linker (green, 384–396), β -sheet subdomain of substrate-binding domain with groove (orange, 397–502), α -helical lid subdomain of substrate binding domain (yellow, 503–602)³⁸. **b**, DnaK in ATP bound open conformation (PDB ID: 4B9Q). **c**, Apo MBP (PDB ID: 2MV0). Proteins are displayed in the same scale to aid visualizing interactions between them.



Extended Data Figure 4 | Predicted peptides binding to DnaK for MBP, luciferase and RepE54. **a**, Peptide-library-trained predictions of DnaK-interacting peptide segments in the unfolded chain of MBP mapped in orange, using the algorithm introduced by Rüdiger *et al.*¹² (PDB ID: 1ANF). N terminus marked green, C terminus marked blue. Surface

representations of predicted peptides are restricted to their backbone atoms as their accessibility is central to canonical peptide binding of DnaK. **b**, Idem for MBP core structure. **c**, Idem for firefly luciferase (PDB ID: 1LCI) **d**, Idem for RepE54 (PDB ID: 1REP) including DNA ligand.



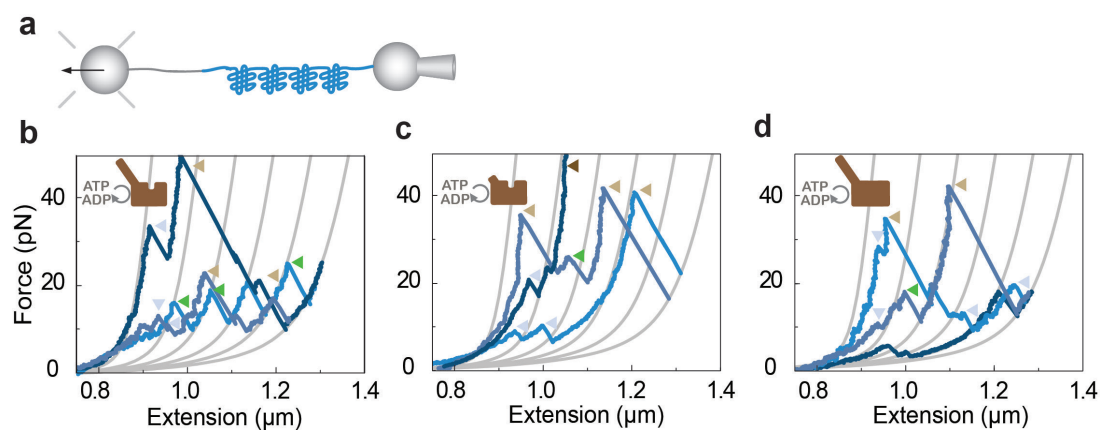
Extended Data Figure 5 | Stabilization of folded structures by DnaK.

a, Schematic diagram of the setup. DnaK and ADP were present in the experimental buffer from the start of the experiment. **b**, In the first stretching curve (blue), only the C-terminal fragment of MBP was unfolded, and then force was immediately reduced to low force preventing further unfolding. After 3 min waiting at low force, subsequent stretching showed resistance to forced unfolding up to 50 pN (orange). Experiments performed in 100 nM DnaK with 1 mM ADP loading buffer. **c**, Force acting on MBP is plotted versus time. Pulls (single stretching–relaxation cycles) on the same MBP molecule are followed by increasing waiting periods at 0 pN, in the presence of 1 μ M DnaK and 1 mM ADP loading buffer. **d**, Force–extension curves of MBP indicate increasing protection of partial folds against force (corresponding to panel c). Earlier pulls are in blue, later in red. **e**, Force on MBP versus time. Identical buffer conditions as panels c and d. **f**, Force–extension curves of MBP indicate the

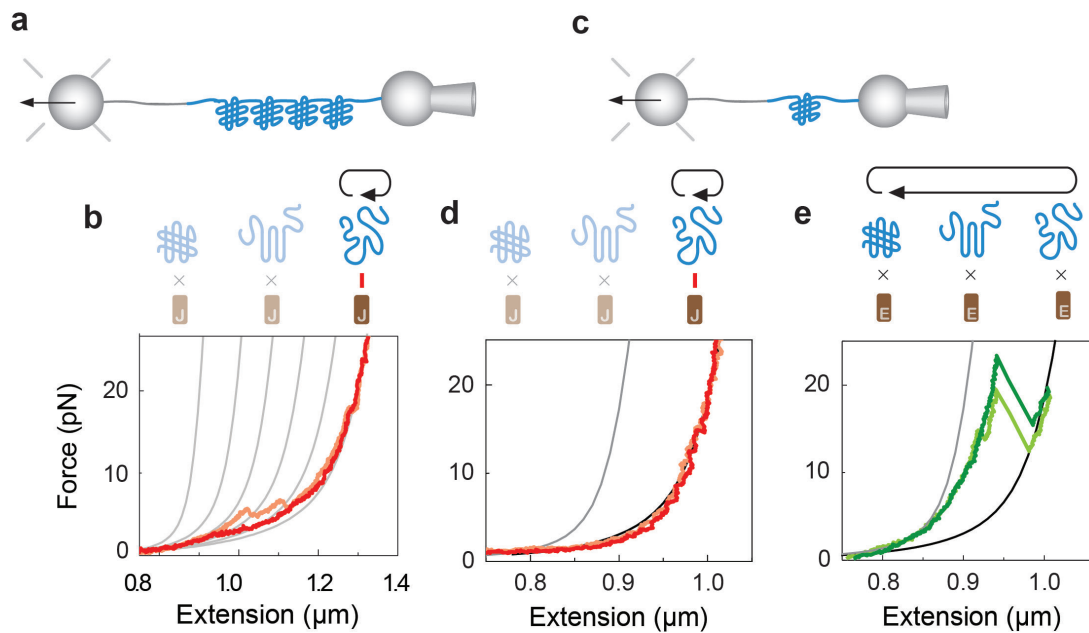
stabilization of compact state of MBP (corresponding to panel e). First pull is highlighted in blue, subsequent pulls in red. **g**, Force–extension curves of MBP in the presence of 100 nM DnaK and 1 mM ADP loading buffer. Blue, first stretching curves on different MBP molecules; orange and red, stretching curves after refolding and 3 min waiting at low force. **h**, **i**, MBP stretching and relaxation experiments in the presence of 1 μ M DnaK with 100 μ M ADP purified by high-performance liquid chromatography (HPLC) were present. First stretching curves are shown in blue, stretching curves denoted in shades of red were acquired after 3 min waiting at low force. **j**, MBP stretching and relaxation in the presence of 100 nM DnaK mutant T199A and 1 mM ATP, which traps DnaK in the ATP state. **k**, Corresponding refolding probability ($n=18$). **l**, Force–extension curves of MBP in the presence of 1 μ M lid-truncated DnaK and 1 mM ADP loading buffer, after refolding and 3 min waiting at low force ($n=13$).

a**b**

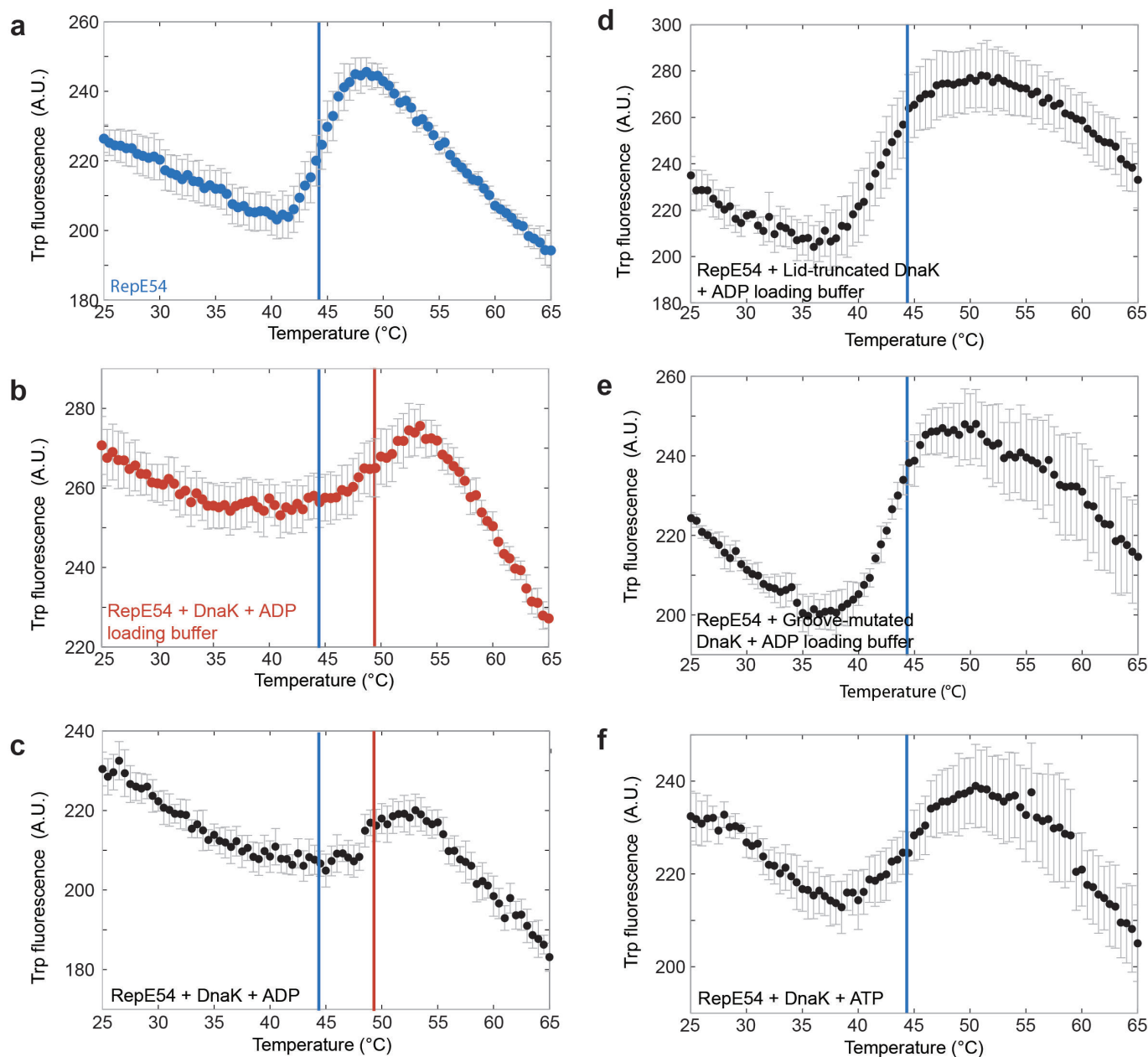
Extended Data Figure 6 | DnaK mutant structures. **a**, DnaK(V436F) (termed groove-mutated) structure model highlighting in red spheres the location of the point mutation hindering access to its groove, resulting in a 38-fold reduction in peptide-binding affinity²⁴ (V436F substitution modelled onto PDB ID: 2KHO). **b**, DnaK(2–538) (termed lid-truncated) structure (lid deletion modelled using 2KHO). The part of the lid that is still present in DnaK(2–538) interacts with the nucleotide binding domain in the ATP-bound open conformation, which might be important for the mechanics of DnaK.



Extended Data Figure 7 | 4MBP refolding in the presence of ATP and wild-type or mutant DnaK. **a**, Diagram of the experiment. **b–d**, Stretching curves of 4MBP in the presence of ATP and wild-type DnaK (**b**), lid-truncated DnaK (**c**), and groove-mutated DnaK (**d**).

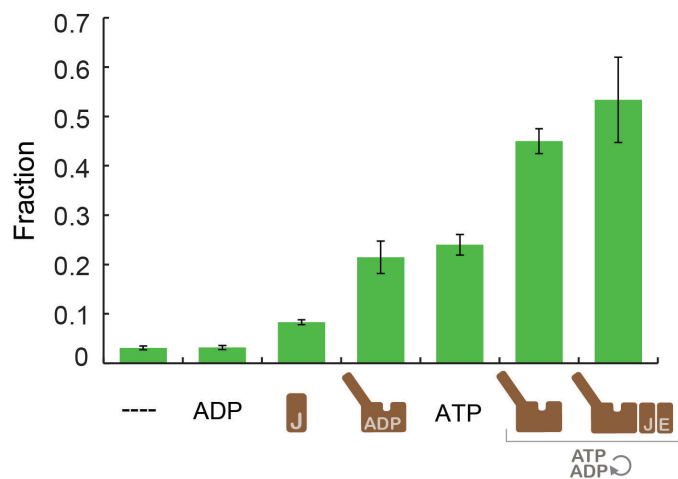


Extended Data Figure 8 | MBP interaction with co-chaperone DnaJ and nucleotide exchange factor GrpE. **a**, Diagram of the 4MBP experiments. **b**, Stretching curves of 4MBP with DnaJ. **c**, Diagram of the 1MBP experiments. **d**, 1MBP stretching in the presence of DnaJ. **e**, 1MBP stretching in the presence of GrpE.



Extended Data Figure 9 | DnaK increases thermal stability of RepE54. **a–f**, Thermal denaturation curves of RepE54 as measured by tryptophan fluorescence in the absence (**a**) or presence (**b**, **c**, **f**) of tryptophan-free DnaK(W102F) with ADP loading buffer (**b**), ADP (**c**), ATP (**f**) or

lid-truncated (**d**) or groove-mutated (**e**) tryptophan-free DnaK with ADP loading buffer. Vertical lines mark the apparent melting points of RepE54 only (blue) and DnaK-stabilized RepE54 (red). Error bars indicate the standard error of the mean over three replicates.



Extended Data Figure 10 | DnaK preserves enzyme activity. Bulk luciferase protection functional assay monitoring luciferase activity at 37°C in the absence or presence of nucleotides and chaperones as indicated. The active fraction of luciferase after 45 min from the start of temperature upshift from 0°C to 37°C is shown. The experiment is performed in triplicates; error bars indicate the standard error of the mean.

Local regulation of gene expression by lncRNA promoters, transcription and splicing

Jesse M. Engreitz^{1,2}, Jenna E. Haines^{1,†}, Elizabeth M. Perez¹, Glen Munson¹, Jenny Chen^{1,2}, Michael Kane¹, Patrick E. McDonel^{1,†}, Mitchell Guttman³ & Eric S. Lander^{1,4,5}

Mammalian genomes are pervasively transcribed^{1,2} to produce thousands of long non-coding RNAs (lncRNAs)^{3,4}. A few of these lncRNAs have been shown to recruit regulatory complexes through RNA–protein interactions to influence the expression of nearby genes^{5–7}, and it has been suggested that many other lncRNAs can also act as local regulators^{8,9}. Such local functions could explain the observation that lncRNA expression is often correlated with the expression of nearby genes^{2,10,11}. However, these correlations have been challenging to dissect¹² and could alternatively result from processes that are not mediated by the lncRNA transcripts themselves. For example, some gene promoters have been proposed to have dual functions as enhancers^{13–16}, and the process of transcription itself may contribute to gene regulation by recruiting activating factors or remodelling nucleosomes^{10,17,18}. Here we use genetic manipulation in mouse cell lines to dissect 12 genomic loci that produce lncRNAs and find that 5 of these loci influence the expression of a neighbouring gene in *cis*. Notably, none of these effects requires the specific lncRNA transcripts themselves and instead involves general processes associated with their production, including enhancer-like activity of gene promoters, the process of transcription, and the splicing of the transcript. Furthermore, such effects are not limited to lncRNA loci: we find that four out of six protein-coding loci also influence the expression of a neighbour. These results demonstrate that cross-talk among neighbouring genes is a prevalent phenomenon that can involve multiple mechanisms and *cis*-regulatory signals, including a role for RNA splice sites. These mechanisms may explain the function and evolution of some genomic loci that produce lncRNAs and broadly contribute to the regulation of both coding and non-coding genes.

We analysed 12 lncRNA loci whose RNA transcripts in mouse embryonic stem cells (mES cells) show preferential localization to the nucleus and span a range of abundance levels (Methods and Extended Data Fig. 1). For each locus, we looked for direct regulatory effects on local gene expression by using a genetic approach based on classical *cis*–*trans* tests (Fig. 1a and Supplementary Note 1). Specifically, we generated clonal cell lines carrying heterozygous knockouts of the promoter (~600–1,000-bp deletions) (Fig. 1b) and compared the expression of nearby genes within 1 Mb on the *cis* and *trans* alleles (that is, on the modified and unmodified homologous chromosomes in the same cells) (Supplementary Note 2). Changes in neighbouring gene expression that involve only the *cis* allele very probably result from direct, local functions of the lncRNA locus, while changes that involve both the *cis* and *trans* alleles probably result as indirect, downstream consequences of the lncRNA acting elsewhere (Supplementary Note 1). We performed genetic modifications in 129/*castaneus* F1 hybrid mES cells that contain a polymorphic site every ~140 bp, enabling us to distinguish the two alleles using RNA sequencing (Fig. 1b, Extended Data Fig. 2 and Supplementary Note 3).

At 5 of these 12 lncRNA loci, promoter knockouts significantly affected the expression of a nearby gene in an allele-specific manner (false discovery rate <10%), including both activating and repressive effects (Fig. 1c, d, Supplementary Note 4 and Extended Data Fig. 3). For each locus, the affected gene was located immediately adjacent to, and within 5–71 kb of, the knocked-out promoter (Fig. 1c and Extended Data Fig. 4). This indicates that a substantial fraction of lncRNA loci influence the expression of a neighbouring gene.

To test whether such effects were specific to lncRNA loci, we deleted the promoters of six protein-coding genes (Extended Data Fig. 1). Surprisingly, knockouts at four of these loci also affected the expression of a neighbour in *cis* (Fig. 1c, d and Extended Data Fig. 5). Thus, both non-coding and coding loci can directly influence local gene expression. These regulatory connections may contribute to the observed correlations in the expression of neighbouring genes, which have been reported both for lncRNAs and for mRNAs^{10,11,19,20}.

Because in these experiments we deleted gene promoters, the mechanisms underlying such *cis* effects could in principle involve (i) DNA regulatory elements in gene promoters^{13–16}; (ii) the process

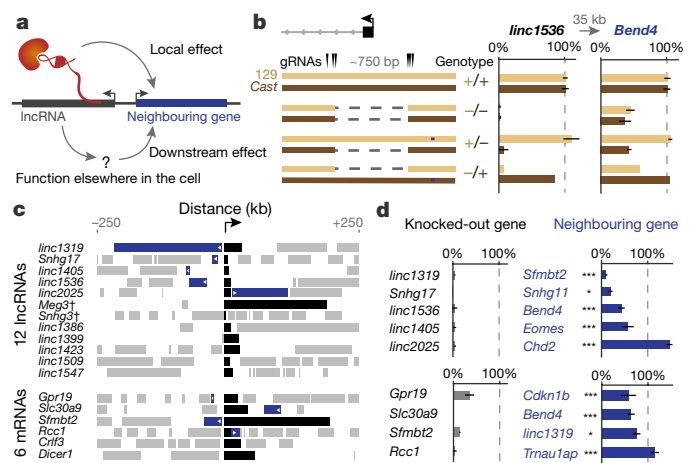


Figure 1 | Many lncRNA and mRNA loci influence the expression of neighbouring genes. **a**, Knocking out a promoter (black) could affect a neighbouring gene (blue) directly (local) or indirectly (downstream). **b**, Knockout of the *linc1536* promoter. Left, genotypes; right, allele-specific RNA expression for 129 and *castaneus* (*Cast*) alleles normalized to 81 control clones (+/+). Error bars, 95% confidence interval for the mean ($n = 2$ for $-/-$, 3 for $+/-$, 1 for $-/+$). **c**, Gene neighbourhoods oriented so each knocked-out gene (black) is transcribed in the positive direction. Blue neighbouring genes show allele-specific changes in expression. †See Supplementary Note 3. **d**, Average RNA expression on promoter knockout compared to wild-type alleles ($n \geq 2$ alleles, see Supplementary Table 1). *FDR < 10%; ***FDR < 0.1%.

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ²Division of Health Sciences and Technology, MIT, Cambridge, Massachusetts 02139, USA. ³Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA. ⁴Department of Biology, MIT, Cambridge, Massachusetts 02139, USA. ⁵Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02114, USA. †Present addresses: Department of Molecular & Cell Biology, University of California Berkeley, Berkeley, California 94720, USA (J.E.H.); University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA (P.E.M.).

of transcription^{10,17,18}; or (iii) the RNA transcripts themselves^{5–9} (Extended Data Fig. 6a). To begin to distinguish among these possible mechanisms, we inserted early polyadenylation signals (pAS) 0.5–3 kb downstream of each transcription start site (TSS) that eliminated the production of most of the RNA while leaving the promoter sequence intact (Fig. 2 and Extended Data Fig. 6b, c, see Methods). We examined four lncRNA loci and two mRNA loci where promoter deletion affected the expression of a neighbouring gene (see Supplementary Note 5).

As one example, we describe the *linc1536* locus, hereafter called *Bendr* (*Bend4*-regulating effects not dependent on the RNA, Fig. 2a). Whereas deleting the *Bendr* promoter reduced the expression of the adjacent *Bend4* gene by 57%, inserting a pAS into the first intron of *Bendr* (~570 bp downstream of the TSS in this ~13 kb locus) had no effect on *Bend4* expression despite eliminating the spliced *Bendr* RNA (Fig. 2b, c). Furthermore, global run-on sequencing (GRO-seq) did not detect any transcriptionally engaged polymerase upstream of the pAS insertion (Fig. 2c and Extended Data Fig. 7a), perhaps because the pAS prevents RNA splicing, which may substantially reduce transcriptional activity in the modified locus^{21,22}. Therefore, *cis* activation of *Bend4* requires neither the mature *Bendr* RNA transcript nor significant *Bendr* transcription. Instead, this effect appears to be mediated by DNA regulatory elements in the ~750 bp knocked-out promoter-proximal region.

In total, at five of the six loci examined with pAS insertions (including three lncRNAs and two mRNAs), DNA regulatory elements in the promoter-proximal sequences appear to be responsible for activating a neighbouring gene (Extended Data Fig. 7b). Although the promoters in these loci would not be classified as 'enhancers' based on H3K4me3/H3K4me1 ratios²³, they are bound by mES cell transcription factors (Extended Data Fig. 7c) and are located in close proximity to their neighbouring target genes (Fig. 1c and Extended Data Fig. 7d, e), suggesting that these promoters may affect local gene expression through mechanisms similar or identical to enhancers^{13,24,25}.

We also identified one locus, *linc1319* (renamed *Blustr*: bivalent locus (*Sfmbt2*) is upregulated by the splicing and transcription of an RNA), where both promoter deletions and pAS insertions substantially reduced the expression of a neighbouring gene, *Sfmbt2*, located 5 kb upstream (Fig. 3a). To dissect the regulatory mechanism, we tested whether the activation of *Sfmbt2* is mediated by a sequence-specific function of the *Blustr* transcript or the process of transcription (by which we mean one or more sequence-independent functions associated with transcription, such as changes in chromatin state or recruitment of co-factors). To test the first possibility, we knocked out each of the three downstream exons and three introns. None of these deletions impaired *Sfmbt2* activation (Fig. 3b, Supplementary Note 6), suggesting that the activation of *Sfmbt2* does not require unique sequences or structures in the *Blustr* transcript itself. To test the second possibility, we engineered pAS insertions at five different locations in the first exon or intron (+40 bp to +15 kb downstream of the TSS) and found that increasing the length of the *Blustr* transcribed region led to increased activation of *Sfmbt2* (Fig. 3b and Extended Data Fig. 8a, b). We note that changing the length of the transcribed region affected the total amount of engaged polymerase in the *Blustr* locus (Fig. 3c). Thus, *Sfmbt2* activation responds to changes in the length/amount of transcriptional activity in the *Blustr* locus but does not appear to require specific sequence elements in the mature *Blustr* transcript (Supplementary Note 7).

Because promoter-proximal splice sites and the process of splicing can enhance transcription—in some cases by as much as 100-fold^{21,22}—we tested whether the splicing of *Blustr* is involved in *Sfmbt2* activation. Upon deleting the 5' splice site of the first intron of *Blustr* (Extended Data Fig. 8c), we observed a 94% reduction in *Blustr* transcription (as assayed by GRO-seq), a 92% reduction in the levels of the mature *Blustr* transcript, and an 85% reduction in *Sfmbt2* expression (Fig. 3b, c and Extended Data Fig. 8a, b), demonstrating that the first 5' splice site of *Blustr* has a critical role in activating *Blustr* and *Sfmbt2* transcription. By contrast, downstream splice sites were dispensable: upon deleting

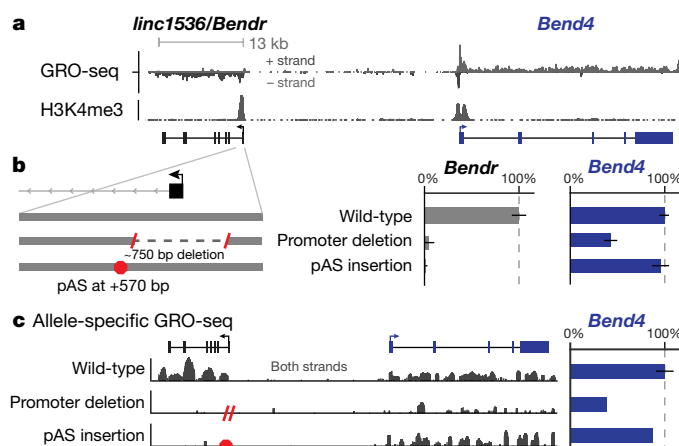


Figure 2 | Enhancer-like function of the *Bendr* promoter.

a, Transcriptionally engaged RNA polymerase (GRO-seq) and H3K4me3 occupancy (chromatin-immunoprecipitation followed by sequencing, ChIP-seq). **b**, poly(A)⁺ RNA expression upon deleting the *Bendr* promoter or inserting a pAS on modified alleles versus controls. Error bars, 95% confidence interval for the mean ($n \geq 2$ alleles, see Supplementary Table 1). **c**, Allele-specific GRO-seq signal for clones carrying the indicated modifications. Both clones are modified on the 129 allele, and only reads specifically mapping that allele are shown. The y axis shows normalized read count. Bar plot quantifies signal at *Bend4*, including seven additional wild-type controls not shown on left.

downstream *Blustr* exons, splicing skipped over the removed exon to the next available 3' splice site (Extended Data Fig. 8d) and *Sfmbt2* expression was unaffected (Fig. 3b).

Together, these data demonstrate that the 5' splice site and the process of transcription in the *Blustr* locus are important for its ability to regulate *Sfmbt2*. This indicates that the *Blustr* RNA is in fact required for *Sfmbt2* activation (splicing involves direct interactions between the spliceosome and the nascent transcript), although this mechanism does not appear to depend on the precise sequence of the RNA beyond the presence of initial splice signals. One possibility is that the 5' splice site promotes transcriptional activity in the *Blustr* locus, which in turn recruits components of the transcriptional machinery that act on the nearby *Sfmbt2* promoter (Fig. 3d, Supplementary Note 7). Consistent with this model, altering transcription or splicing in the *Blustr* locus led to changes in chromatin state at the *Sfmbt2* promoter (including reductions in H3K4me3 and spreading of H3K27me3) and reduced occupancy of engaged RNA polymerase in the paused position just downstream of the *Sfmbt2* TSS (Extended Data Fig. 8b, e, f). Thus, changes in *Blustr* transcription and splicing may affect *Sfmbt2* expression in part by altering chromatin state and RNA polymerase occupancy at the *Sfmbt2* promoter (Fig. 3d and Supplementary Note 7).

In summary, genetic dissection of 12 lncRNA loci and 6 mRNA loci found that 9 loci (50%) regulate the expression of a neighbouring gene (Extended Data Fig. 9). In most of these loci, including *Bendr*, local effects are mediated by enhancer-like functions of DNA elements in promoters. In one locus, *Blustr*, the processes of transcription and splicing also contribute to *cis*-regulatory functions, perhaps by increasing the local concentration of transcription-associated factors. We did not identify any lncRNA loci in which local effects are mediated by sequence-specific functions of the lncRNA transcript. Because there exist thousands of other loci that fit our selection criteria, we expect that similar mechanisms broadly contribute to gene regulation in many loci (Supplementary Note 8).

The frequent cross-talk between neighbouring genes observed in our study indicates that gene loci can encode multiple independent categories of functions. Category I involves functions of the RNA product: mRNAs provide a template for protein synthesis, and some non-coding transcripts (for example, XIST) act as functional lncRNAs. Category II involves the effects of transcription-related

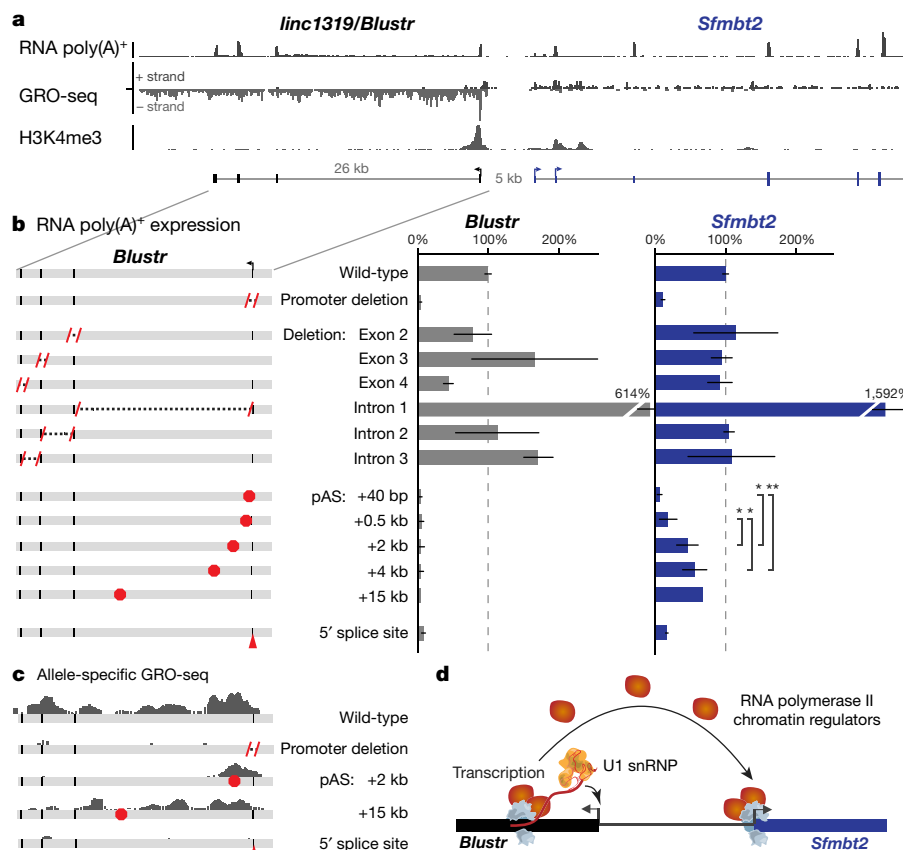


Figure 3 | Transcription and splicing of *Blustr* activates *Sfmbt2* expression. **a**, Poly(A)⁺ RNA-seq, GRO-seq, and H3K4me3 ChIP-seq in the *Blustr* locus. *Sfmbt2* has two alternative TSSs. **b**, Poly(A)⁺ RNA expression on modified alleles compared to controls (arrows). Error bars, 95% confidence interval for the mean ($n \geq 2$ alleles, except for pAS

+15 kb where $n = 1$, see Supplementary Table 1). *Sfmbt2* pAS comparisons: two-sided t -test. * $P < 0.05$; ** $P < 0.01$. **c**, Allele-specific GRO-seq signal for clones carrying indicated modifications. Only reads mapping to the modified allele are shown (*castaneus* for pAS +2 kb; 129 for others). **d**, Model for how transcription in the *Blustr* locus activates *Sfmbt2*.

processes—including mechanisms mediated by promoters, transcription, and splicing—on the regulation of other nearby genes.

The fact that many lncRNA loci have category II functions does not necessarily mean that they do not also have category I functions, and we note that our experiments do not rule out the possibility that the lncRNAs dissected in this study have RNA-mediated functions other than on local gene regulation. However, the prevalence of category II functions suggests a model for the evolutionary origins of some lncRNAs. In loci where a promoter acts as an enhancer, RNA transcripts may arise as non-functional by-products¹⁶. In loci where co-transcriptional processes have *cis*-regulatory functions, the nascent transcripts might contribute through mechanisms such as splicing that require little RNA-sequence specificity. These possibilities are particularly intriguing in light of the patterns of evolutionary conservation of lncRNA loci^{26–28}. For example, although most lncRNA transcripts expressed in mES cells are not conserved (no RNA detected in syntenic loci in other mammals, see Methods), the promoters in some of these loci correspond to conserved DNA sequences that have an enhancer chromatin signature in human ES cells (Fig. 4, Extended Data Fig. 10 and Supplementary Note 9). These sequences may have conserved functional roles as *cis*-regulatory elements, rather than as lncRNA promoters. Thus, mechanisms associated with *cis* functions by promoters, transcription, and/or RNA processing may contribute to the functions and evolution of an important subset of non-coding loci in mammalian genomes (Extended Data Fig. 10c).

Beyond the implications for lncRNAs, these *cis*-regulatory connections between neighbouring genes occur in both protein-coding and non-coding loci and thus appear to represent a fundamental property of mammalian gene regulatory networks. The properties of these

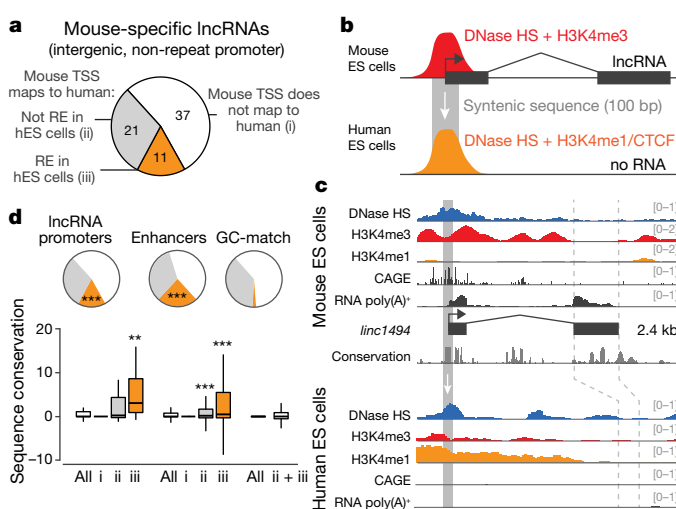


Figure 4 | Evolutionary conservation of mES cell lncRNAs and their promoters. **a**, Classification of a subset of lncRNAs expressed in mES cells (see Supplementary Note 9, Methods). **b**, Eleven lncRNAs have promoters whose syntenic sequence corresponds to putative DNA regulatory elements (REs) marked by DNase I hypersensitivity (HS) in human ES cells. **c**, Example: *linc1494*. **d**, Enhancers and lncRNA promoters are significantly enriched for corresponding to human regulatory elements (pie chart, *** $P < 1 \times 10^{-10}$, χ^2 test versus GC-matched random regions) and show elevated sequence conservation compared to GC-matched regions (bar plot, ** $P < 0.01$; *** $P < 0.001$, Mann–Whitney U -test versus ii + iii).

cis-regulatory connections—including mechanisms for specificity and the potential for cooperative dynamics of gene activation—represent key areas for future investigation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 April; accepted 10 October 2016.

Published online 26 October 2016.

- Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
- Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
- Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
- Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
- Lee, J. T. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* **23**, 1831–1842 (2009).
- Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
- Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
- Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
- Guil, S. & Esteller, M. *Cis*-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.* **19**, 1068–1075 (2012).
- Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. *Nat. Cell Biol.* **10**, 1106–1113 (2008).
- Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
- Bassett, A. R. *et al.* Considerations when investigating lncRNA function *in vivo*. *eLife* **3**, e03058 (2014).
- Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
- Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
- Yin, Y. *et al.* Opposing roles for the lncRNA *haunt* and its genomic locus in regulating *HOXA* gene activation during embryonic stem cell differentiation. *Cell Stem Cell* **16**, 504–516 (2015).
- Paralkar, V. R. *et al.* Unlinking an lncRNA from its associated *cis* element. *Mol. Cell* **62**, 104–110 (2016).
- Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature* **429**, 571–574 (2004).
- Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference—a crash course. *Trends Genet.* **21**, 339–345 (2005).
- Purmann, A. *et al.* Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. *Genomics* **89**, 580–587 (2007).
- Kosak, S. T. *et al.* Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol.* **5**, e309 (2007).
- Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E. & Palmiter, R. D. Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl Acad. Sci. USA* **85**, 836–840 (1988).
- Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929–933 (2001).
- Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
- Andersson, R., Sandelin, A. & Danko, C. G. A unified architecture of transcriptional regulatory elements. *Trends Genet.* **31**, 426–433 (2015).
- Kim, T.-K. & Shiekhattar, R. Architectural and functional commonalities between enhancers and promoters. *Cell* **162**, 948–959 (2015).
- Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
- Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Reports* **11**, 1110–1122 (2015).
- Chen, J. *et al.* Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* **17**, 19 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Grossman, J. Rinn, M. Yassour, P. Sharp, L. Boyer, M. Ray, C. Fulco, M. Munschauer, T. Wang and N. Friedman for discussions; A. Goren and Broad Technology Labs for ChIP; J. Lis, D. Mahat and A. Shishkin for technical advice and reagents; and J. Flannick for computational tools. J.M.E. is supported by the Fannie and John Hertz Foundation and the National Defense Science and Engineering Graduate Fellowship. M.G. is supported the NIH Director's Early Independence Award (DP5OD012190), the Edward Mallinckrodt Foundation, the Sontag Foundation, and the Searle Scholars Program. Work in the Lander Laboratory is supported by the Broad Institute.

Author Contributions J.M.E., M.G. and E.S.L. conceived and designed the study. J.M.E., J.E.H., G.M., M.K. and P.E.M. developed knockout protocols and performed genetic manipulations. E.M.P. and J.M.E. performed all other experiments. J.M.E. developed computational tools and analysed data. J.M.E. and J.C. performed evolutionary analysis. J.M.E. and E.S.L. wrote the manuscript with input from all authors. E.S.L. supervised the work and obtained funding.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.S.L. (eric@broadinstitute.org).

METHODS

Cell lines and cell culture. F1 hybrid 129/*castaneus* female mouse embryonic stem cells (gift from K. Plath) or V6.5 male mouse embryonic stem cells (gift from A. Meissner) were cultured in serum-free N2B27-based medium (250 ml neurobasal media (Gibco), 250 ml DMEM/F12 (Gibco), 5 ml 100× N2 supplement (Gibco), 5 ml 50× B27 supplement (Gibco), 5 ml 200 mM L-glutamine (Gibco), 3.6 µl 2-mercaptoethanol, 50 µg human leukaemia initiation factor (5×10^5 units, EMD Millipore), 7.4 µg progesterone, 10 mg bovine insulin (Sigma), 350 µl 7.5% BSA fraction V (Gibco), supplemented with MEK inhibitor PD0325901 (50 µl 10 mM, SelleckChem), and GSK3b inhibitor CHIR99021 (150 µl 10 mM, SelleckChem)). Prior to plating cells, tissue culture dishes were pre-treated with PBS + 0.2% gelatin (Sigma) and 1.75 µg ml⁻¹ laminin (Sigma) for 2–10 h at 37°C. At each passage, cells were trypsinized for 3–5 min in TVP solution (0.025% trypsin, 1% chicken serum (Sigma), and 1 mM EDTA in PBS pH 7.4) at room temperature. Cells tested negative for mycoplasma contamination and were authenticated by comparing polymorphisms to 129S1 and *castaneus* genomes.

Cellular fractionation. To estimate the relative abundance of lncRNAs in different cellular compartments, we performed cellular fractionation to isolate chromatin-associated, soluble nuclear, and cytoplasmic fractions essentially as described²⁹. In brief, we first lysed 5 million cells in 200 µl cold cell lysis buffer (10 mM Tris-HCl pH 7.5, 0.05% IGEPAL CA-630, 150 mM NaCl), incubating on ice for 5 min. We layered the cell lysate over 2.5 volumes of chilled sucrose cushion (24% sucrose in cell lysis buffer) and centrifuged at 15,000g for 10 min. The supernatant from this spin became the cytoplasmic fraction. After washing the pellet of nuclei with PBS (pH 7.5) + 1 mM EDTA, we resuspended the pellet in 100 µl of cold glycerol buffer (20 mM Tris-HCl pH 7.5, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, 0.125 mM PMSE, 50% glycerol) by gently flicking the tube. We added 100 µl of cold nuclei lysis buffer (10 mM HEPES pH 7.5, 1 mM DTT, 7.5 MgCl₂, 0.2 mM EDTA, 0.3 M NaCl, 1 M urea, 1% IGEPAL CA-630), then vortexed for four seconds. After 2 min on ice, we spun the nuclear lysate at 15,000g for 2 min. This supernatant was collected as the soluble nuclear (nucleoplasm) fraction. We rinsed the remaining pellet (chromatin fraction) in PBS + 1 mM EDTA, then resuspended the chromatin in 300 µl chromatin DNase buffer (20 mM Tris-HCl pH 7.5, 50 mM KCl, 4 mM MgCl₂, 0.5 mM CaCl₂, 2 mM TCEP, 0.5 mM PMSE, 0.4% sodium deoxycholate, 1% IGEPAL CA-630, 0.1% N-lauroylsarcosine) plus 15 µl murine RNase inhibitor (NEB) and 30 µl TURBO DNase (Ambion). The DNase digestion proceeded for 20 min at 37°C and was halted by adding 10 mM EDTA and 5 mM EGTA. Protein was digested with proteinase K for 1 h at 37°C. RNA was isolated using Zymo RNA Concentrator-25 columns (two columns for the cytoplasmic fraction). With this method, nuclear-associated endoplasmic reticulum is known to fractionate with the nucleoplasm²⁹, and we observed that nucleolar RNAs fractionated with chromatin (data not shown). From each cellular fraction, we sequenced total RNA and polyadenylated RNA (selected using oligo d(T)₂₅ magnetic beads, NEB) using a strand-specific RNA-sequencing protocol for Illumina instruments described previously³⁰.

Selection criteria for knocked-out lncRNAs. We selected lncRNA loci initially identified and defined by a chromatin signature of H3K4me3 at promoters and H3K36me3 through gene bodies³. We further required that lncRNAs selected for knockout analysis have TSSs, as defined by cap analysis of gene expression (CAGE), located >5 kb from other genes (for epigenomic annotation of each locus, see <http://pubs.broadinstitute.org/neighbor-genes/>). To prioritize intergenic lncRNA loci that may regulate local gene expression, we focused on lncRNAs that have subcellular localization biased towards the nucleus versus the cytoplasm (Extended Data Fig. 1). We performed cellular fractionation experiments in V6.5 male mES cells as described above and sequenced RNA from chromatin-associated, soluble nuclear, and cytoplasmic fractions (GEO Accession GSE80262). We calculated a relative nuclear-to-cytoplasmic ratio (chromatin RPKM plus soluble nuclear RPKM divided by cytoplasmic RPKM) and focused on lncRNAs with ratios above the median (1.5): these lncRNAs are preferentially localized to the nucleus compared to other lncRNAs and mRNAs. We selected nuclear-biased lncRNAs that span a range of abundance levels (Extended Data Fig. 1). We also included some lncRNAs that are conserved across mammalian evolution (*Snhg3*, *Snhg17*, *Meg3*, and *linc2025*).

Selection criteria for knocked out mRNAs. We selected six mRNAs for promoter knockouts based on the following criteria. We knocked out two mRNAs that are moderately expressed and are not expected to be essential for mES cell growth (*Dicer1* and *Crlf3*). We knocked out two mRNAs that are located adjacent to knocked-out lncRNAs (*Sfmbt2* and *Rcc1*), in order to look for reciprocal regulatory effects between the lncRNA and the affected mRNA. We knocked out two mRNAs that are located adjacent to a gene that is itself adjacent to a lncRNA (*Gpr19* and *Slc30a9*), in order to determine whether affected genes are specifically responsive to lncRNA promoters or are generally responsive to other promoters in the locus.

Similar to the lncRNAs selected, the TSSs of these selected mRNAs are located >5 kb from other genes.

CRISPR sgRNA design. To design single-guide RNAs (sgRNAs), we built custom software to calculate a specificity score (based on potential off-target sites using the algorithm described at <http://crispr.mit.edu> (see ref. 31)) and an efficacy score (based on a sequence model for sgRNA efficiency as previously described³²) for each 20-nt targeting sequence. We removed guides with specificity scores <20 or efficacy scores >0.7. To avoid T-rich sequences that result in premature termination of Pol III-mediated sgRNA transcription, we removed guides with more than one T in the four bases closest to the seed region, guides with more than three consecutive Ts, and guides with more than eight Ts total. We removed guides with homopolymer stretches of five or more bases and guides with GC content <20% or >90%. We removed guides that overlapped a known 129/*castaneus* SNP³³. Within a given region, we typically chose the three remaining guides with the highest specificity scores. The sequences of all sgRNAs used in this study are listed in Supplementary Table 2.

Promoter deletion guide placement. To knock out a lncRNA or mRNA promoter, we chose 2–3 sgRNAs located in windows 300–500 bp upstream and downstream of the TSS, leading to deletions of approximately 600–1,000 bp surrounding the TSS. We adjusted the precise deletion boundaries outward if we could not successfully design guides in these regions (for example, because they were located in repetitive sequences). We note that we often found that the wild-type alleles in heterozygous knockouts were affected by scars from repair of sgRNA double-stranded breaks. Accordingly, we adjusted the bounds if necessary to cut outside of the exons of the mRNA or lncRNA and thus avoid damaging the exonic sequences on the wild-type alleles in heterozygous knockouts. We note that the presence of these scars (and their lack of allele-specific effects on the expression of neighbouring genes) indicate that the *cis* effects observed upon deleting promoters are not merely a result of CRISPR-mediated cutting and subsequent DNA repair.

Genetic deletions with CRISPR/Cas9. To delete specific sequences, we co-transfected 100 ng of Cas9-expressing plasmids (PX330-NoGuide), 300 ng of a pool of sgRNA-expressing plasmids (pZB-Sg3), and 100 ng of a plasmid expressing EGFP and a puromycin selectable marker from a CAG promoter (pS-pp7-GFPiP). To create PX330-NoGuide, we modified PX330 (gift from F. Zhang, Addgene plasmid #44230 (ref. 34)) to remove the sgRNA expression cassette. To generate pZB-Sg3, we cloned a human U6 promoter and optimized sgRNA scaffold sequence³⁵ into a minimal vector with an ampicillin-selectable marker and a ColE1 replication origin. We transfected batches of 250,000 mouse embryonic stem cells using the Neon Transfection System (Invitrogen), using one pulse of 40 ms at 1,200 V and plated two batches of cells (500,000 total) into a 96-well plate in 200 µl media. As an internal control for each set of transfections, we performed a transfection using four guides with no predicted target sites in the mouse genome.

We verified efficient transfection by examining GFP expression after 24 h. To select for transfected cells, we replaced the media 24 h after transfection with 200 µl 2i + 1 µg ml⁻¹ puromycin. One day later, we split the cells into a 10-cm plate with 8 ml of 0.5 µg ml⁻¹ puromycin. One day later, we replaced the media with 10 ml of 2i with no puromycin. We allowed cells to grow for 7–8 days, replacing the media every 2–3 days. We hand-picked 88 individual colonies and 8 control colonies for each transfection in 5 µl media, added 20 µl of TVP for ~10–20 min at 37°C to dissociate the colonies, and then split the colonies into two identical plates. We grew the cells in these plates for 4–5 days. We harvested one of the plates for DNA and RNA extraction by removing most of the media and adding 3.5 × volume buffer RLT (Qiagen) and froze the other plate for later recovery in Freezing Media (2i media + 10% fetal bovine serum + 10% DMSO).

Genotyping by PCR and sequencing. To genotype each promoter knockout, we extracted genomic DNA and performed PCR using primers spanning the deleted sequence. We genotyped each clone by running the PCR products on agarose gels and comparing PCR amplicon sizes to predicted wild-type and deletion band sizes. We confirmed the sequences of wild-type and deletion bands by Sanger sequencing or high-throughput sequencing through barcoded amplicon sequencing on an Illumina MiSeq (see Supplementary Table 2). Where possible, we used known polymorphic sites from 129S1 and *castaneus* genomes³³ to determine the haplotype-resolved genotype of each clone. Based on the genotyping data, we nominated clones for RNA sequencing. We eliminated clones showing evidence of polyclonal or subclonal mutations, or complex mutations such as inversion or duplication of the genomic sequence between the sgRNAs. The sequences of all genotyping primers are listed in Supplementary Table 2.

RNA sequencing libraries. We generated RNA sequencing libraries as previously described^{30,36}, with some modifications for high sample throughput. We isolated RNA from harvested mES cells using RNeasy 96 columns. We enriched for poly(A)⁺ RNA using oligo d(T)₂₅ magnetic beads (NEB) and eluted in 18 µl H₂O.

We fragmented RNA to an average of ~150 nt by adding 2 µl Ambion fragmentation buffer and incubating at 70 °C for exactly 2.5 min. After transferring quickly to ice, we added 40 µl of a master mix containing 12 µl 5× FNK buffer (50 mM Tris-HCl pH 7.5, 5 mM MgCl₂, 0.6 mM CaCl₂, 50 mM KCl, 10 mM DTT, 0.01% Triton X-100), 1 µl Murine RNase Inhibitor (NEB), 3 µl FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), 3 µl T4 Polynucleotide Kinase (NEB), and 1 µl TURBO DNase (Life Technologies). We incubated this reaction at 37 °C for 30 min, then cleaned the reaction with MyOne SILANE magnetic beads³⁷ and eluted in 6 µl of H₂O.

We proceeded with the library preparation as previously described³⁰, with one additional modification. To simplify the library preparation for many samples, we added unique sample barcodes (8 nt) during the first adaptor ligation³⁶. We used 12 pools each with 4 barcodes in order to mitigate differences in the efficiency of ligation for different adaptor sequences. Following the first adaptor ligation, we pooled 12 samples together, including up to 9 clones corresponding to a single target gene as well as 3 control clones, during the first 70% ethanol wash of the SILANE-bead purification. We performed an extra SILANE purification using the same beads to remove excess adaptor and then proceeded with reverse transcription.

Hybrid selection of RNA sequencing libraries. To measure allele-specific expression for hundreds of genes in a cost-effective manner, we developed a hybrid selection strategy to enrich for allele-informative reads at target genes (Extended Data Fig. 2). We designed oligo pools to capture allele-informative sequences in the ~1,600 RNAs located in the genome within 1 Mb of one of the knockout targets. These target RNAs were divided into two independent pools: #140820 and #141203. We used RefSeq RNA annotations for mRNAs and our custom annotations for most lncRNAs. We identified SNPs that would distinguish the 129S1 and *castaneus* genomes³³. We designed 120-bp capture oligos in the vicinity of each 129/*castaneus* polymorphic site, tiling every 15 bp across either 600 bp (pool #140820) or 240 bp (pool #141203) centred on the SNP. We included probes targeting both alleles to minimize differences in capture efficiency between the two alleles. We filtered capture probe sequences as previously described³⁷. We included up to 10 oligos per targeted RNA, duplicating probes where necessary to include the sequences corresponding to each allele. Empirically, this probe design strategy in combination with the protocol described below enabled assessing allele-specific expression for 84% (611 of 731) of the targeted expressed genes in mES cells (RPKM ≥ 2) at a sequencing depth of <5 million reads per sample. Target genes and oligo sequences for these pools are listed in Supplementary Table 3.

We synthesized pools of 12,000 capture oligos using CustomArray technology. Oligos in each pool were flanked by unique primers (Left primer sequence: CTTCTACGAGCAGTTTGCC; right primer sequence: AGTTTACGCATTACGGGCAC). After one round of PCR to add a T7 promoter (GGATTCTAATACGACTCACTATAGGG), we generated biotinylated RNA probes as described previously³⁸, adding in 20% Biotin-16-UTP (Roche) and 20% Biotin-14-CTP (Life Technologies) to the *in vitro* transcription reactions. We generated RNA probes targeting both strands by incorporating the T7 promoter into either side of the PCR product and performing two separate *in vitro* transcription reactions per oligo pool.

To capture the allele-informative regions, we pooled the final, barcoded RNA sequencing libraries from all samples in the batch and performed a modified version of solution hybrid selection³⁹. We first combined 500 ng dsDNA library pool with 1 nmol of Illumina P5 and P7 primer mix in 21 µl total. We denatured this mix at 94 °C for 10 min and transferred immediately to ice. We added 7.5 µl 20× SSPE, 0.5 µl Murine RNase Inhibitor (NEB), and 1 µl of 500 ng µl⁻¹ biotinylated RNA probe, for a total volume of 30 µl. We set up at least two reactions per 10 libraries, including at least one reaction with each strand of probes. We incubated the hybridization reaction at 65 °C for 24–48 h. For each capture sample, we washed 30 µl Streptavidin C1 MyOne magnetic beads (Invitrogen) in 5× SSPE and aliquoted them into PCR tubes. After removing the wash from the beads, we added the hybridization reaction and mixed to resuspend the beads. We captured the biotinylated probes by shaking at 65 °C for 20 min. We washed the beads twice in 150 µl low stringency wash buffer (1× SSPE, 0.1% SDS, 1% NP-40, 4 M urea) at 62 °C for 3–4 min, and twice in 150 µl high stringency wash buffer (0.1× SSPE, 0.1% SDS, 1% NP-40, 4 M urea). To elute, we removed the final wash and resuspended beads in 10 µl 100 mM NaOH and heated to 70 °C for 10 min. To complete the elution, we added 1 µl 1 M acetic acid and 14 µl NLS elution buffer (20 mM Tris-HCl pH 7.5, 10 mM EDTA, 2% N-lauroylsarcosine, 2.5 mM TCEP) and heated to 94 °C for 4 min. While hot, we placed samples on magnet, removed eluate, and then placed the eluate on ice for at least 30 s. We cleaned the eluates with 20 µl MyOne SILANE magnetic beads as described³⁷, using 75 µl RLT and 61 µl 100% ethanol for the initial precipitation. We eluted in 23 µl H₂O, and used this as input for a 50 µl NEBNext High Fidelity PCR reaction using 500 pmol of the P5 and P7 Illumina primers each (98 °C for 30 s; 13 cycles of 98 °C for 15 s, 68 °C for 30 s, 72 °C for 30 s;

72 °C for 2 min, 4 °C hold). We cleaned the PCR reaction twice with 1× volume Agencourt Ampure XP magnetic beads and eluted in 20 µl H₂O.

Allele-specific gene expression measurements from RNA sequencing. We sequenced RNA libraries on an Illumina HiSeq 2500 (Read 1: 38 cycles; Read 2: 30 cycles; Index: 8 cycles). The first read includes the 8-nt barcode added during the first adaptor ligation (see above). Following processing to separate samples based on the inline barcodes, we filtered out sequencing reads that aligned to highly abundant RNA transcripts, including ribosomal RNAs, snRNAs, and repetitive elements, as defined by RefSeq and RepeatMasker. A FASTA file containing these sequences is available at the Gene Expression Omnibus (GSE55914).

We developed a computational pipeline to estimate allele-specific expression from RNA-sequencing data. We created two separate reference files for the 129S1 and *castaneus* haplotypes, starting with the mm9 genome build and layering on SNPs based on whole-genome sequencing of each of the two mouse strains³³. We aligned RNA-sequencing data separately to each of the two haplotypes using TopHat (version 2.0.8). We combined the results of the two alignments using PySuspenders⁴⁰, which identifies reads that map specifically to one or the other allele and splits them into separate BAM files. We discarded duplicate reads and reads with MAPQ < 30. After generating separate BAM files containing the reads mapping to each allele, we counted reads that mapped to each RefSeq transcript (including both spliced and unspliced isoforms) using Scripture⁴¹ and calculated 'allelic expression ratios' for each gene (counts from 129 allele divided by total counts from both 129 and *castaneus* alleles). The distribution of allelic expression ratios for all active genes in mES cells was centred on 0.5, indicating that on average each gene is expressed equally from the 129 and *castaneus* alleles (Extended Data Fig. 2b). This indicates that there is not systematic bias in our mapping procedure towards one allele or the other.

RNA-seq data analysis. We processed RNA-sequencing data sets in batches corresponding to sets of libraries made on the same day with the same hybrid selection probe pool. We removed samples with fewer than 100,000 non-repetitive, unique, allele-informative reads. For within-batch quality control, we performed hierarchical clustering on all samples by their allelic expression ratios and removed the 2–5% of outlier samples, which were largely comprised of clones that showed monoallelic expression from the X chromosome.

Assessment of gene knockout by expression analysis. The PCR genotyping procedure described above provided putative genotypes for the cell clones. We confirmed the genotype of cells by analysing the allele-specific expression of the knocked-out gene in each clone. We required that clones show >80% reduction of expression of the knocked out gene on the appropriate allele in order to include the clone in downstream analysis. Incomplete reduction of expression in some cases appeared to result from use of alternative TSSs that were not included in the deleted sequence. In other cases, incomplete reduction of expression appeared to result from subclonal genetic mosaicism within the cell line, which probably resulted from deletions that occurred after several cell divisions, leading to genetic differences between individual cells in a colony. For further analysis, we focused on gene loci where we obtained at least two heterozygous knockout clones.

Barplots for allele-specific expression data. Barplots that depict allele-specific RNA expression or GRO-seq transcription on modified alleles compared to controls (for example, Fig. 1d) were calculated as follows. For each modified allele, allele-specific measurements were normalized to the corresponding alleles in wild-type clones (for example, values for *castaneus* knockout alleles were divided by the mean of unmodified *castaneus* alleles in wild-type clones). We performed the same calculation for unmodified alleles in wild-type clones to create a null distribution. For modified alleles, we further scaled these values by dividing by the mean of the wild-type alleles in heterozygous knockout clones. The value of each bar represents a mean of these normalized measurements.

Identifying significant changes in allele-specific expression. In developing a statistical approach to identify local, *cis* effects of these genetic manipulations, we sought to distinguish local effects of the genetic deletion from downstream effects that result as a consequence of either lncRNA/mRNA functions elsewhere in the cell, off-target effects, or biological/technical variation between clonal cell lines (Supplementary Note 1). Our power to detect these effects varies between different measured genes (owing to their level of expression and availability of SNPs) and between different knockout targets (owing to differences in the numbers of knockout clones analysed).

To account for these two variables, we developed a statistical approach to empirically estimate the false discovery rate of allele-specific changes in the expression neighbouring genes using hundreds of genes on other chromosomes as controls. For each gene in the neighbourhood of one of our promoter deletions, we calculated three statistics: (i) a *t*-test statistic comparing the average change in expression for each of the knockout alleles (including both heterozygous and homozygous knockout clones), normalized to the expression of the gene on the wild-type allele of the heterozygous clones; (ii) a *z*-score statistic comparing the

expression of the knockout allele in heterozygous clones to the expression of the wild-type allele in the same clone; and (iii) a *t*-test statistic comparing the heterozygotes to the wild-type control clones using the allelic expression ratio after applying a variance-stabilizing transformation (arcsin of the square root of the allelic expression ratio). For a given gene, only samples with at least 20 allele-informative reads were considered, in order to enable accurate estimates of allele-specific expression. These three tests differ in whether they incorporate information from homozygous clones and how they normalize between knockout and wild-type alleles. We required that a gene perform significantly in each of the three tests in order to regard the gene as significant, as described below. We note that each underlying measure was approximately normally distributed, with some apparent outliers across hundreds of control clones; we conservatively included these outliers in calculating each test statistic. We examined differences in variation between knockout and control alleles with Levene's test. For estimates of the variance of distributions presented in figures, see Supplementary Table 1.

Because the distributions are only approximately normal, we assessed the significance of each of these gene-level statistics by permutation, sampling other cell lines from the same experimental batch and randomly assigning them as heterozygous or homozygous knockout clones to match the distribution of genotypes of the real samples. We calculated an empirical false discovery rate for the sum of these permutation ranks, testing each of the neighbouring genes and using all of the genes on other chromosomes as the background model. Neighbouring genes with FDR <10%, a transformed allelic expression ratio >0.03, and an effect size of >10% in heterozygotes were considered significant.

No statistical methods were used to predetermine sample size, but we generated as many knockout clones as possible. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Transcriptional read-through for *Meg3* and *Snhg3*. Promoter knockouts of *Meg3* and *Snhg3* led to reductions in one or more downstream genes oriented in the same direction as the knockout target gene. We attributed these changes to transcriptional read-through based on the following evidence (Supplementary Note 4 and Extended Data Fig. 3). For both *Meg3* and *Snhg3*, we observed evidence for transcription continuing past the annotated 3' end of the knockout target, through intergenic regions, and into the downstream gene (as assayed by RNA sequencing of chromatin-associated RNA). For the *Meg3* locus, we did not observe H3K4me3 or CAGE reads at the 5' ends of *Rian* and *Mirg* (downstream of *Meg3*), indicating that they are not expressed from their own promoters. In the *Snhg3* locus, the downstream affected gene (*Rcc1*) is in fact expressed from its own promoter, but we found evidence for reads splicing from just downstream of *Snhg3* into the first splice acceptor of *Rcc1*, indicating that at least some fraction of *Rcc1* transcripts begin at the *Snhg3* promoter.

Insertion of polyadenylation signals. To halt transcription, we initially attempted to use a short 49-bp synthetic polyadenylation signal (spA) sequence⁴² to minimize the amount of genomic sequence added (Extended Data Fig. 6b). For a given gene, we designed a guide 0.5–3 kb downstream of the transcription start site. We designed 200-nt ssDNA oligos including the spA sequence flanked by 75- and 76-bp homologous arms, centred on the sgRNA cut site (~4 bp upstream of the PAM sequence), and ordered these as ultramers from Integrated DNA Technologies (Supplementary Table 2). To knock in polyadenylation signals, we transfected 100 ng PX330-NoGuide, 100 ng pZB, 100 ng pS-pP7-GFPiP, and 100–200 ng of donor ssDNA oligo and followed the selection procedure described for the promoter knockouts. To genotype these insertions, we used a combination of PCR and high-throughput amplicon sequencing as described above. We identified clones that had heterozygous insertions of the full 49-bp spA sequence on one allele; we typically observed that the other allele had a short insertion or deletion, consistent with non-homologous end joining (NHEJ)-mediated repair. This short pAS sequence (spA) succeeded in halting the transcription of three RNAs: *Blustr* (pAS at +40 bp and +0.5 kb in Fig. 3), *Gpr19*, and *Bendr*. However, for other genes, transcription was unaffected despite pAS knock-in, consistent with the location-dependent efficiency previously observed for this pAS sequence⁴².

Accordingly, we built a larger construct containing three polyadenylation signals (p3PA, Extended Data Fig. 6c). The structure of this construct upon insertion into the genome through homologous recombination is as follows: spA–EFS promoter–Puromycin resistance gene IRES thymidine kinase–WPRESV40 pAS–PGK pAS (p3PA–Puro–iTk). We co-transfected 300 ng of this construct with 100 ng of pZB and 100 ng of PX330-NoGuide, waited three days, and then selected for cells with integrations with 1 μ g ml^{−1} puromycin for one week. We picked individual colonies and used PCR to genotype clones, using primers spanning the insertion junctions. We sequenced these PCR products to determine the allele of insertion. Following genotyping, we expanded clonal cell lines and transfected them with PX330 and a pool of four sgRNAs to delete the selection cassette, leaving behind three tandem pASs. Following selection with 2 μ g ml^{−1} ganciclovir, we again picked individual

colonies, used PCR to confirm loss of the cassette, and sequenced RNA from multiple clones. PCR primer sequences for cloning homology arms and genotyping p3PA insertions are listed in Supplementary Table 2.

Knockouts of *Blustr* exons and introns. To delete each exon and intron of *Blustr*, we transfected cells with pools of guides as described for the promoter deletions, using two guides on each side. We assessed the genotype of clonal cell lines as described above for promoter deletions. To confirm exon knockout from RNA sequencing data, we examined SNPs in each of the exons. Upon knockout of exon 2, for example, we observed loss of RNA sequencing reads mapping to exon 2, while reads mapping to other exons were still present. We also identified reads spanning a new splice junction between exon 1 and exon 3, further confirming that exon 2 was removed from the mature transcript. For bar plots in Fig. 3 measuring *Blustr* expression, the values represent the normalized read counts of the remaining exons that were not deleted in that experiment. To confirm intron knockout, we used PCR primers spanning the deletion junction and sequenced the resulting PCR products. We note that the intron knockouts, by design, do not affect the sequence of the spliced *Blustr* RNA.

5' splice site knockout. To knock out the 5' splice site of *Blustr*, we co-transfected mES cells as described above, using a single sgRNA pZB plasmid and 200 ng of ssDNA oligonucleotide donor for homologous recombination (Extended Data Fig. 8c). The oligo was ordered as an ultramer from Integrated DNA Technologies (Supplementary Table 2). We genotyped these insertions through amplicon sequencing using an Illumina MiSeq (primers in Supplementary Table 2).

Transcriptional activity with GRO-Seq. We used precision run-on sequencing (PRO-seq)⁴³, a variant of global run-on sequencing⁴⁴, to map transcriptionally engaged RNA polymerase for a subset of clones. Clones for PRO-seq (as well as ChIP-seq and assays for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)) were chosen from among the recoverable knockout cell lines with a preference for clones with homozygous knockouts or knockouts on the 129 allele only. We performed PRO-seq as previously described⁴⁵, with modifications. We harvested 10 million mES cells by scraping, washing in cold PBS, and spinning at 330g for 3 min. The cell pellet was resuspended in 1 ml cold douncing buffer (10 mM Tris-HCl pH 7.4, 300 mM sucrose, 3 mM CaCl₂, 2 mM MgCl₂, 0.1% (v/v) Triton X-100, and 0.5 mM DTT) per 1 million cells. The cells were incubated on ice in the cold room for 5 min and dounced 25 times. The nuclei were pelleted at 500g for 2 min, washed twice in 5 ml douncing buffer, and centrifuged at 500g for 2 min. The nuclei were then gently resuspended in 100 μ l of cold storage buffer (10 mM Tris-HCl, pH 8.0, 25% (v/v) glycerol, 5 mM MgAc₂, 0.1 mM EDTA, and 0.5 mM DTT), immediately flash frozen, and stored at −80 °C until use.

A 28 μ l 2 \times Nuclear Run-On (NRO) mix was prepared as follows: 1 M Tris-HCl, pH 8.0, 1 M MgCl₂, 2 M KCl, and 0.1 M DTT. 5 μ l of 1 mM Biotin-11-CTP (Perkin Elmer), 1 μ l of 0.05 mM CTP, 2.5 μ l of 2 mM ATP, 2.5 μ l of 2 mM GTP, 2.5 μ l of 2 mM UTP (Sigma Aldrich), 6.5 μ l of nuclease free water, and 2 μ l of SUPERaseIn (Ambion) were added to the 2 \times NRO mix and mixed well before the addition of 50 μ l of 2% NLS. The NRO reaction mix was mixed well and preheated to 37 °C. 100 μ l of NRO mix was added to 100 μ l of nuclei in storage buffer. The reaction was mixed gently by pipetting and incubated at 37 °C for 3 min, mixing halfway through. To halt the reaction 500 μ l of TRIzol LS (Thermo Fisher) was added, mixed well, and incubated at room temperature for 5 min. RNA was isolated through a chloroform extraction and ethanol precipitation, and resuspended in 20 μ l of H₂O. The RNA was heat denatured at 65 °C for 40 s and fragmented on ice for 10 min with 5 μ l of 1N NaOH. To stop the reaction, 5 μ l of 1 M acetic acid and 20 μ l of 1 M Tris-HCl, pH 7.4 were added. To remove unincorporated biotinylated nucleotides, the sample was passed through a P-30 exchange column (BioRad). 1 μ l of RNase inhibitor was added to the ~50 μ l of RNA and the first biotin enrichment was then performed.

Each biotin enrichment was performed as follows. To prepare the Streptavidin M280 Beads (Invitrogen) for biotin enrichment, 100 μ l of beads were taken per sample and washed once in 0.1N NaOH with 50 mM NaCl and twice in 100 mM NaCl. Beads were resuspended in 160 μ l of binding buffer (10 mM Tris-HCl, pH 7.4, 300 mM NaCl, and 0.1% (v/v) Triton X-100). To each sample an equal volume of Streptavidin M280 beads was added, mixed, and incubated on a rotator for 20 min at room temperature. The beads were magnetically separated and washed twice in 500 μ l of ice cold high salt wash buffer (50 mM Tris-HCl, pH 7.4, 2 M NaCl, and 0.5% (v/v) Triton X-100), twice in 500 μ l of binding buffer, and once in 500 μ l of low salt wash buffer (50 mM Tris-HCl, pH 7.4 and 0.1% (v/v) Triton X-100). To harvest the RNA, 300 μ l of TRIzol (Thermo Fisher) was added to the beads, vortexed for 20 s, and incubated at room temperature for 3 min. 60 μ l of chloroform was added and mixture was incubated at room temperature for 3 min. The samples were centrifuged at 14,000g for 5 min at 4 °C. The aqueous phase was collected and transferred to a new tube; the remaining organic phase was removed from the beads. The TRIzol extraction was then repeated as above and the two aqueous phases were combined. RNA was purified with a chloroform

extraction and ethanol precipitation, and resuspended in nuclease-free water. RNA sequencing libraries were then prepared as described above, except that SILANE clean-ups were replaced with streptavidin-biotin capture enrichments until after reverse transcription (a total of three enrichments).

We sequenced PRO-seq libraries to a depth of ~10 million 30-bp paired-end reads. To analyse the data, we mapped and processed the RNA sequencing data as described above, including aligning individually to the 129 and *castaneus* genomes. Figures showing 'Allele-specific GRO-seq' depict coverage for reads that uniquely map to the specific allele indicated in the figure. To assess the relative read density in the promoter-proximal region and gene body of *Sfmbt2*, we counted reads in the 2 kb region downstream of the first *Sfmbt2* TSS and in the remainder of the gene body⁴⁶. We calculated the pause index as the ratio of these two quantities, normalized to total read count. We noticed that different PRO-seq libraries had subtle biases in the relative fraction of reads aligning to the TSS versus the gene body, leading to slightly offset distributions of pause indices across all genes, and so we corrected for these biases in each library by normalizing TSS and gene body RPKMs to the median of the ~5,000 genes with coverage across all samples.

Chromatin accessibility with ATAC-seq. Libraries were generated as previously described⁴⁷ using 50,000 mES cells. We generated duplicate ATAC-seq libraries for each clonal cell line examined and sequenced each to a depth of ~40 million 30-bp paired end reads. We aligned paired-end DNA sequencing reads using Bowtie2 (ref. 48) to each of the 129 and *castaneus* genomes with the following parameters: `-met-stder--maxins 1000`, removed duplicate reads using Picard (<http://picard.sourceforge.net>), and filtered to uniquely aligning reads using samtools (MAPQ < 30, <https://github.com/samtools/samtools>). For plotting normalized read coverage at the *Blustr* and *Sfmbt2* promoters, we combined data from the two biological replicates (two independent measures of the same cell line) and connected paired-end reads to generate fragments. Fragment coverage was normalized by the total number of uniquely mapping reads.

Chromatin immunoprecipitation. ChIP-seq for H3K4me3 and H3K27me3 was performed using monoclonal antibodies as previously described⁴⁹. Sequencing data was analysed as for ATAC-seq described above.

Validation of allele-specific RNA expression with ddPCR. To validate our RNA-seq based measurements of allele-specific expression, we used a quantitative allele-specific PCR assay to verify measurements for *Blustr* and *Sfmbt2*. We isolated RNA from harvested mES cells using RNeasy 96 columns and performed a DNase treatment followed by reverse transcription of 500 ng of RNA (total reaction volume 20 µl). We performed droplet digital PCR (ddPCR) using Bio-Rad Custom ddPCR Assays that involve qPCR primers flanking a polymorphic site and two allele-specific fluorescent probes. For *Blustr*: left primer sequence: GACAAATACTCCCTTCAACA; right primer sequence: GAACAGTTTGTCTGCTGCC; probe sequence: TAAGTGAGGTGAACCTCAAG (129 allele, FAM) or AGTGAGGCGAAGCTTCAAG (*castaneus*, HEX). For *Sfmbt2*: left primer sequence: TGTAAGTTTGCCTGATCTC; right primer sequence: TCTAATGTACCTCAGCCC; probe sequence: TTTCTATGAGCAGTTCAAC (129 allele, FAM) or TCCTATGAACCGTTCAGC (*castaneus*, HEX). ddPCR was done with 2.2 µl of cDNA, 11 µl of Supremix (BioRad), 1.1 µl of each probe, and 7.7 µl of water per reaction followed by droplet generation. PCR was performed as follows: 95 °C for 10 min; cycling at 94 °C for 30 s and 55 °C for 1 min for a total of 40 cycles; and 98 °C for 10 min. Readout was done using the QX200 Droplet Reader and QuantaSoft Software (BioRad) to determine the total number of droplets containing each allele. We calculated allelic expression ratios from these values and compared it to values generated through RNA-sequencing and hybrid selection of the same RNA samples (Extended Data Fig. 2d, e).

External ChIP-seq, RNA-seq, and DNase HS data. We used the following data from ENCODE⁵⁰: H3K4me3, H3K4me1, H3K27ac, and CTCF ChIP-seq in mES cells (ES-Bruce4); DNase hypersensitivity sequencing in mES cells (E14); H3K4me3, H3K4me1, and CTCF ChIP-seq and DNase HS data in H1-hES cells; and RNA-sequencing data in H1-hES cells (nuclear poly(A)⁺, nuclear total). To assess transcription factor binding to mRNA and lncRNA promoters (Extended Data Fig. 7c), we examined mES cell ChIP-seq peaks available from Kagay et al. at the Gene Expression Omnibus (GSE22562)⁵¹.

DNA purification for examining proximity contacts. To examine the proximity contacts of the *linc1405* locus, we used the RAP-DNA protocol, which we initially developed in order to map RNA localization to chromatin, to capture *linc1405* DNA³⁷. In brief, we cross-linked live cells to fix endogenous chromatin complexes, then purified a target DNA region using a pool of oligonucleotides targeting the *linc1405* locus (Supplementary Table 3). Here, we used probes that are the same strand as the *linc1405* RNA—in this way, we specifically capture the *linc1405* DNA and do not directly capture the *linc1405* RNA itself. We mapped the 3D proximity contacts of the *linc1405* locus through high-throughput sequencing of co-purified DNA and calculated the normalized enrichment to an input DNA library in 1-kb windows (Extended Data Fig. 7e). Annotations for topologically

associated domains (TADs) were downloaded from the Ren Laboratory (<http://chromosome.sdsc.edu/mouse/hi-c/download.html>)⁵².

lncRNA transcript annotations. For evolutionary conservation analysis, we used lncRNA annotations and isoforms previously defined based on RNA sequencing in mouse embryonic stem cells, combining annotations generated with multiple methods (Scripture⁴¹ and slinky²⁸). We filtered the combined list using slinky²⁸ to eliminate transcripts predicted to encode proteins or micropeptides by UCSC, transcripts that partially align to protein-coding genes (for example, pseudogenes or incomplete reconstructions), and species-specific coding gene duplications. Subsequently we performed several manual curation steps. We examined each isoform using a combination of long-read RNA-sequencing data, total chromatin-associated RNA sequencing data, cap analysis of gene expression (CAGE) data, and poly(A)⁺ 3'-end sequencing data from mES cells^{28,30,41,53}. We eliminated transcripts that appeared to result from an extended 3' UTR of an upstream protein-coding transcript. Because the precise 5' ends of transcripts are imprecisely assigned based on RNA-sequencing data alone, we re-assigned 5' ends (TSSs) using a sliding-window approach to find the 10-bp window with the highest number of same-strand CAGE reads within 300-bp of the initial calculated TSS. We additionally manually curated the TSS of each lncRNA, some of which were incorrectly assigned by more than 300 bp, based on CAGE and H3K4me3 ChIP-seq data, and eliminated any where we could not identify the TSS (for example, due to an unmappable sequence or very low abundance).

Analysis of lncRNA and promoter conservation. To categorize lncRNAs by their conservation properties and promoter locations, we examined a set of 307 lncRNAs expressed in mES cells as described above. We assessed the conservation of each lncRNA through a two-step approach. We first used slinky to look in syntenic locations for evidence of lncRNA transcripts in deep poly(A)⁺ RNA-seq of rat, chimp, and human induced pluripotent stem cells (iPSCs)²⁸. lncRNAs called 'conserved' by this first filter have substantial evidence based on RNA-seq that allows for independent reconstruction of the transcript in one or more of these other organisms. We categorized the remaining lncRNAs by the location of their TSS: 71 lncRNAs originate within 500-bp of an mRNA TSS on the opposite strand (divergent); 59 lncRNAs originate within the long-terminal repeats (LTRs) of endogenous retroelements; and 79 lncRNAs have their promoters in intergenic regions that do not overlap with LTRs and do not emerge from a bidirectional mRNA promoter (henceforth, 'intergenic').

Because some conserved lncRNAs might be expressed at too low a level to assemble a transcript *de novo* in a given species, we examined more closely the 79 intergenic lncRNAs that were called 'mouse-specific' in the initial slinky analysis. We applied a second, more stringent threshold to remove lncRNAs misclassified as mouse-specific due to low abundance. For each intergenic lncRNA locus, we used liftOver⁵⁴ to map the 10 bp surrounding the mouse TSS (mm9) to the human genome (hg19) (minMatch = 0.1, UCSC chain). 37 of these transcripts did not lift over at this step, and thus were considered mouse-specific. For the 42 that did lift over, we examined the syntenic region for evidence of poly(A)⁺ RNA-seq data from human iPSCs²⁸ or poly(A)⁺ nuclear-fraction RNA-seq from hES cells (−100 to +900 bp relative to the TSS), or for evidence of poly(A)⁺ nuclear-fraction or whole-cell CAGE from hES cells (−250 to +250 bp relative to the TSS), and removed from consideration any lncRNAs that showed evidence for RNA-seq or CAGE above a certain threshold. We chose this threshold based on a set of random intergenic regions, which were matched to the set of intergenic mouse-specific lncRNAs based on GC content. We eliminated from consideration the ten lncRNAs that showed RNA-seq or CAGE signals greater than the 90th percentile of random regions, corresponding to approximately two CAGE or RNA-seq reads in the windows described above. These ten lncRNAs were added to the 'conserved' section of the pie chart in Fig. 4a. Several of these ten lncRNAs correspond to substantially shortened, single-exon poly(A)⁺ transcripts that show minimal overlap with the syntenic exons in mouse; although a majority of the exonic sequence of these transcripts are not in fact conserved between human and mouse, we excluded these from consideration as putative mouse-specific lncRNAs.

For the purposes of examining the conservation properties of these intergenic mouse-specific lncRNAs, we defined a matched set of 'enhancer' elements. We first generated a list of regulatory elements in mES cells using the DNase hotspots called by ENCODE-UW in ES-E14 cells. As an estimate of the activity of each element, we calculated the density of H3K27ac reads in the region. From the set of intergenic elements that did not overlap a promoter, lncRNA promoter, or LTR, we selected a random subset matched to the intergenic lncRNA promoters for H3K27ac density (binned by 10 reads per bp) and distance to the TSS of the closest active gene (binned by 5 kb). We call these elements 'enhancers' because they are marked by DNase hypersensitivity and H3K27ac but do not overlap a known gene promoter.

We compared the sequence conservation and functional conservation of three classes of elements: intergenic mouse-specific lncRNAs, matched intergenic

enhancer elements, and GC-matched random intergenic elements. First, we computed the rate at which each set maps to human sequence. We centred each element and used liftOver ($-\text{minMatch} = 0.1$) to identify the syntenic region in the human genome. Elements that did not lift over at this step correspond to the white segment of the pie charts in Fig. 4 (iii, did not map). For elements that did lift over to human, we next defined the subset that map to putative regulatory elements in human. We examined a 500-bp window centred on the lifted over region and counted reads in hES cell DNase-seq data from ENCODE. We defined regions showing DNase HS scores higher than 95% of the mappable random intergenic regions as putative DNA regulatory elements. We note that these random intergenic regions include some enhancers; they are matched to lncRNA promoters for GC content, and thus frequently correspond to regulatory elements (which are GC-rich) that happen to be active in hES cells. For both intergenic mouse-specific lncRNAs and enhancers, $\sim 33\%$ of elements corresponded to putative DNA regulatory elements in human (Fig. 4d), representing a ~ 6.6 -fold enrichment versus the random intergenic controls. To compare sequence conservation of these classes of elements, we calculated the average SiPhy score⁵⁵ across each 500-bp region surrounding the mouse TSS or the centre of the enhancer element, using the 29 mammals alignment from the mouse perspective⁵⁶. We used a two-sided Mann–Whitney *U*-test to look for changes in the distributions of SiPhy scores to the set of mappable random intergenic regions (Fig. 4d: random ii + iii).

Impact of expression level on conservation analysis. Although the set of intergenic mES cell lncRNAs examined above does not show any significant evidence for poly(A)⁺ RNA in the syntenic locus in human, some of these transcripts may not be detected in human and yet still be truly conserved. These transcripts might be misclassified as mouse-specific lncRNAs for several reasons, including: (i) low expression level in hES cells and iPS cells such that the lncRNA, by chance, is not detected based on the depth of sequencing data available; or (ii) the lncRNA is not expressed in hES cells or iPS cells, but is expressed in a different human cell type and thus may have a conserved function.

To estimate the false positives resulting from these and other scenarios, we examined the properties of a set of 853 conserved mRNAs matched to the intergenic mouse-specific lncRNAs based on expression in mES cells. We counted the frequency at which these mRNAs would be called ‘not conserved’ by the same procedures described above: we applied the nuclear poly(A)⁺ CAGE and RNA-seq filters to eliminate transcripts that show detectable transcription in the 1-kb region near the TSS. While 87% of the intergenic lncRNAs described above passed these filters (and thus appeared to be mouse-specific), only 22% of the expression-matched mRNAs passed; this indicates that the set of 69 mouse-specific intergenic lncRNAs are approximately 3.9-fold enriched for human elements that are not transcribed in hES cells. Thus, the mouse-specific lncRNAs defined above appear to consist largely of transcripts that are not conserved.

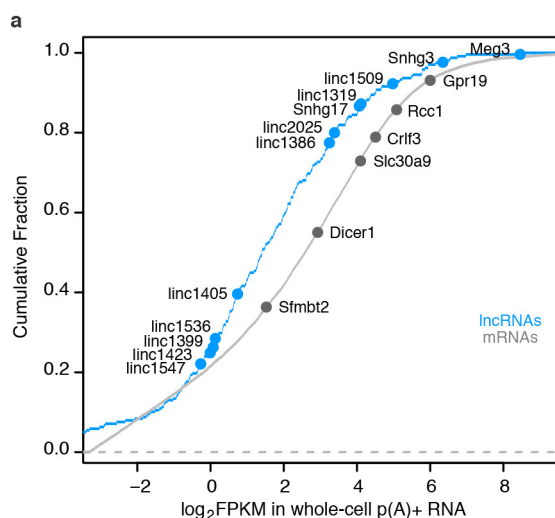
We performed the following additional analyses to ensure the robustness of our conclusions regarding the existence of lncRNAs that evolved from ancestral regulatory elements. First, we examined the conservation of the first 5' splice sites of this set of lncRNAs. In 7 of these 11 loci, the GT dinucleotide in the first 5' splice site is not conserved, suggesting that a similar spliced transcript cannot be produced from this locus. Second, we re-performed the entire conservation analysis focusing on the 50% of mES cell intergenic lncRNAs with the highest expression levels; these lncRNAs are less likely to be missed in hES cells due to low abundance. We also adjusted our poly(A)⁺ RNA and CAGE filters to require a complete absence of reads in the corresponding regions in hES cells and iPS cells. Using these filters, 79% of the intergenic lncRNAs are not detectably expressed in human cells, representing a ~ 12 -fold enrichment over mRNAs matched for expression level. Therefore we are confident that most of these lncRNAs are correctly classified as mouse-specific. Of the 30 intergenic lncRNAs called mouse-specific by this more conservative analysis, 5 do indeed correspond to putative DNA regulatory elements, including *linc1494* (Fig. 4c), representing a > 8 -fold enrichment versus GC-matched random sequences ($\chi^2 P < 1 \times 10^{-10}$). Thus, our conclusions that some lncRNAs appear to evolve from ancestral regulatory elements are robust even with stringent thresholds.

Software for data analysis and graphical plots. We used the following software for data analysis and graphical plots: R Bioconductor (version 3.0)⁵⁷, Gviz (version 1.10.11), gplots (version 2.17.0), GenomicRanges (version 1.18.4)⁵⁸, rtracklayer (version 1.26.3)⁵⁹, BEDTools⁶⁰, Integrative Genomics Viewer (version 2.3.26)⁶¹, and vcftools (version 0.1.12)⁶².

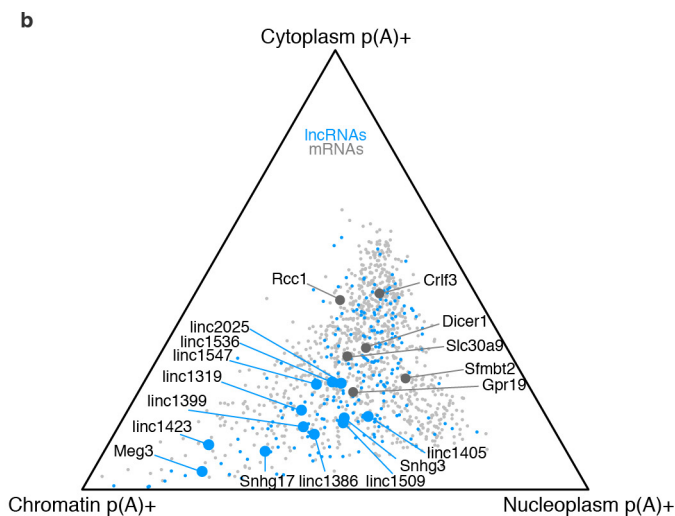
Data availability. Sequencing data for this study is available at the Gene Expression Omnibus (GSE80262 and GSE85798), and additional visualizations of the data are available at <http://pubs.broadinstitute.org/neighboring-genes/>.

Code availability. Code for the analyses described in this paper is available from the authors upon request.

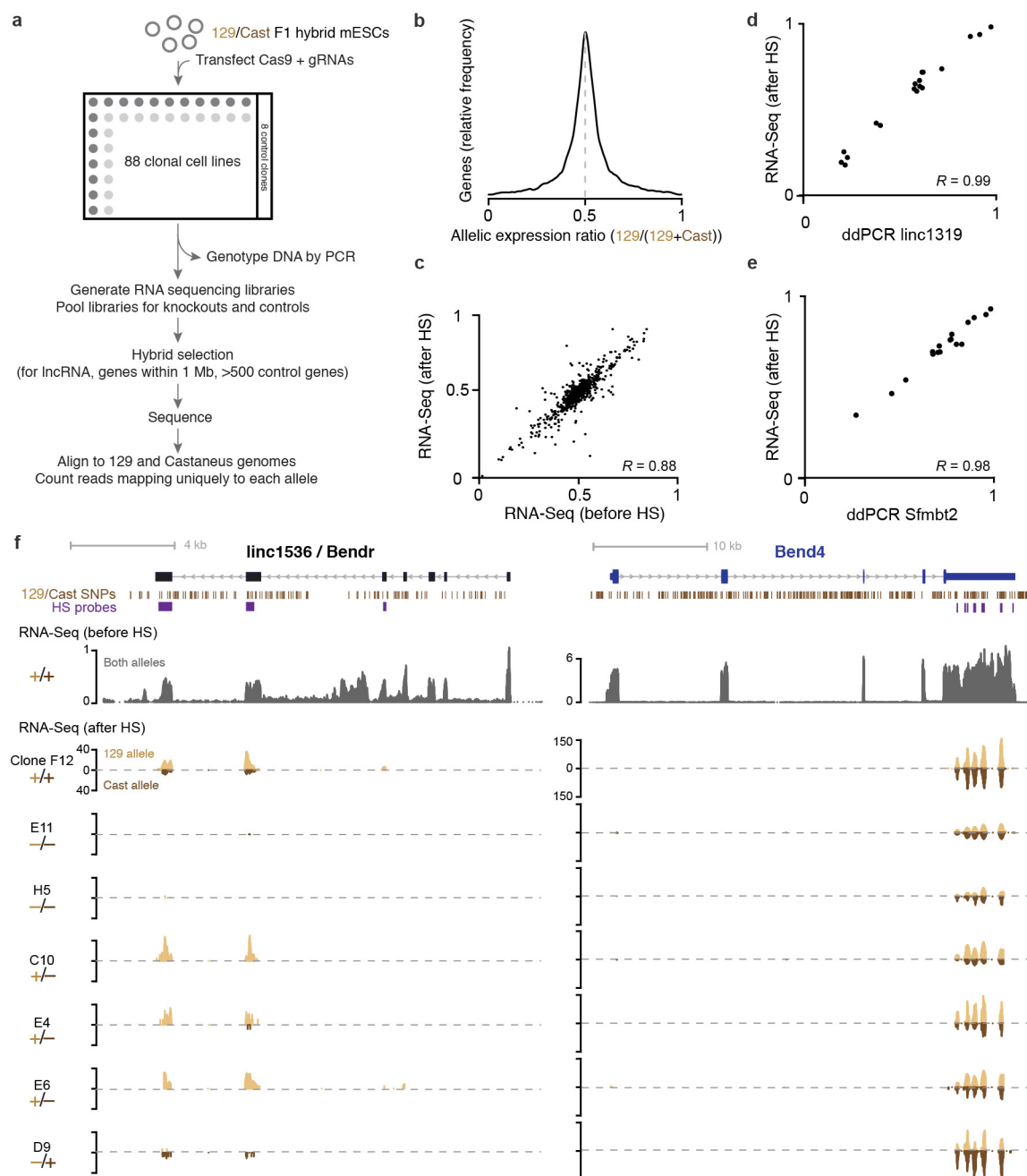
29. Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
30. Engreitz, J. M. *et al.* RNA–RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* **159**, 188–199 (2014).
31. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
32. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
33. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
34. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
35. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
36. Shishkin, A. A. *et al.* Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* **12**, 323–325 (2015).
37. Engreitz, J., Lander, E. S. & Guttman, M. RNA antisense purification (RAP) for mapping RNA interactions with chromatin. *Methods Mol. Biol.* **1262**, 183–197 (2015).
38. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
39. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
40. Huang, S., Holt, J., Kao, C.-Y., McMillan, L. & Wang, W. A novel multi-alignment pipeline for high-throughput sequencing data. *Database (Oxford)* **2014**, bau057 (2014).
41. Guttman, M. *et al.* *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28**, 503–510 (2010).
42. Levitt, N., Briggs, D., Gil, A. & Proudfoot, N. J. Definition of an efficient synthetic poly(A) site. *Genes Dev.* **3**, 1019–1025 (1989).
43. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
44. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
45. Mahat, D. B. *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protocols* **11**, 1455–1476 (2016).
46. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
47. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
49. Busby, M. *et al.* Systematic comparison of monoclonal versus polyclonal antibodies for mapping histone modifications by ChIP-seq. Preprint at <http://dx.doi.org/10.1101/054387> (2016).
50. Mouse ENCODE Consortium *et al.* An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13**, 418 (2012).
51. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
52. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
53. Fort, A. *et al.* Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* **46**, 558–566 (2014).
54. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
55. Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, i54–i62 (2009).
56. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
57. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
58. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
59. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
60. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
61. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
62. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).



Extended Data Figure 1 | Expression and subcellular localization of knocked-out lncRNAs and mRNAs. **a**, Expression of lncRNAs and mRNAs in F1 129/*castaneus* female mES cells, reported in fragments per kilobase per million (FPKM) in whole-cell poly(A)⁺ RNA-seq. Cumulative fraction is plotted for all mRNAs expressed in mES cells. Large dots represent transcripts whose promoters we deleted in this study. lncRNAs and mRNAs span a >20-fold range of abundance levels. **b**, Relative

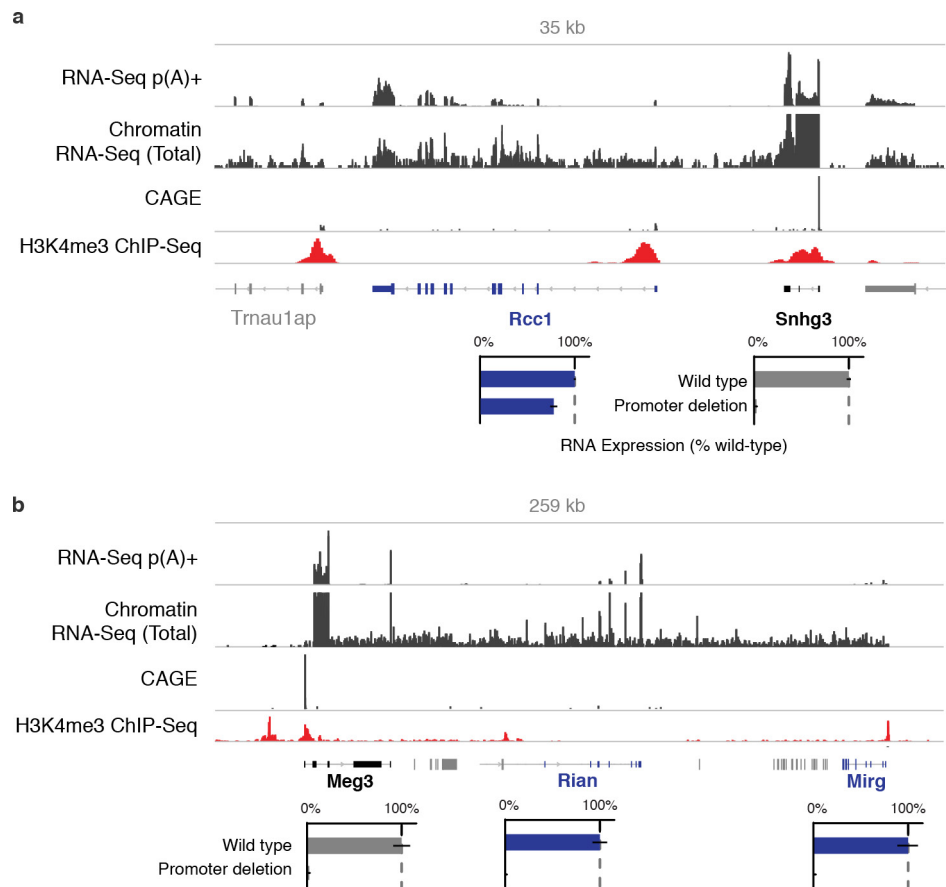


subcellular localization of lncRNAs and mRNAs. We sequenced poly(A)⁺ RNA from chromatin, soluble nuclear, and cytoplasmic fractions (see Methods) and plotted the relative abundance of mature transcripts in each fraction. We selected lncRNAs that showed localization biased towards the nuclear fractions relative to most mRNAs. For comparison, we plotted 1,000 randomly selected mRNAs (light grey).



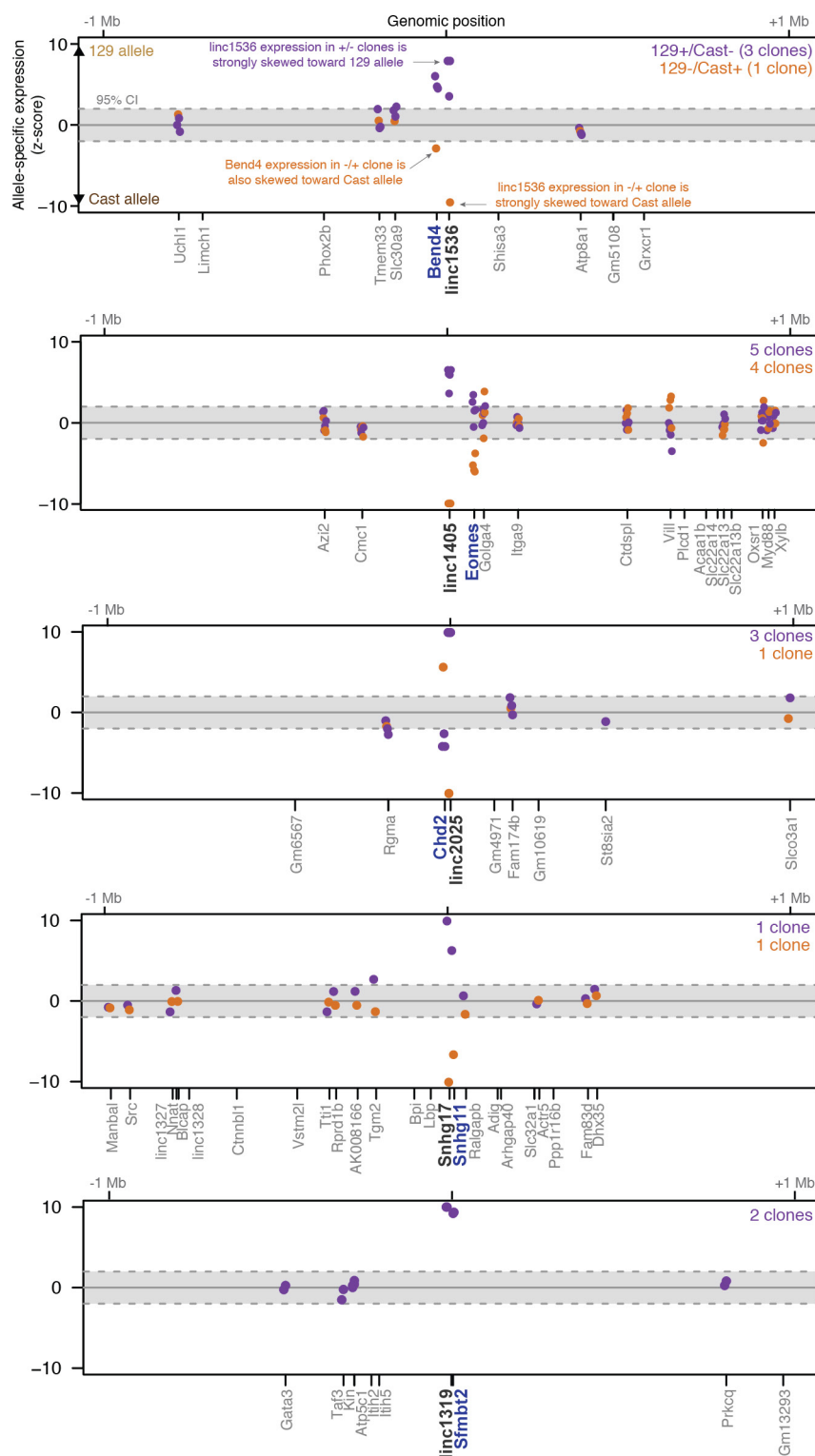
Extended Data Figure 2 | Generation of knockout clones and measurement of allele-specific RNA expression. **a**, Overview of knockout and measurement protocol. **b**, Distribution of allelic expression ratios (number of informative reads mapping to 129S1 allele divided by the number mapping to either the 129S1 or the *castaneus* allele) across active genes in mES cells. **c**, Scatter plot of allelic expression ratios for genes with RPKM ≥ 2 that have more than 100 allele-informative reads across all libraries. Allelic expression ratios are consistent in RNA sequencing data before and after hybrid selection (HS). **d**, **e**, Allelic expression ratios as measured by two independent methods for *Blustr* (**d**) and *Sfmbt2* (**e**) expression in 15 clonal cell lines containing genetic modifications in

the *Blustr* locus. Each dot represents the mean of two ddPCR technical replicates (x axis) and the value from one RNA-seq technical replicate (y axis). **f**, Example locus showing hybrid selection strategy and RNA-seq coverage for cell lines with the indicated genotype for deletion of the *Bendr* promoter. The y axis scales represent normalized read counts and are the same for all hybrid selection tracks. The absolute level of expression for any given gene varies among clonal cell lines; throughout this work, we instead consider the relative level of expression between the two alleles in heterozygous knockout cells. For similar plots of each gene studied, see <http://pubs.broadinstitute.org/neighboring-genes/>.



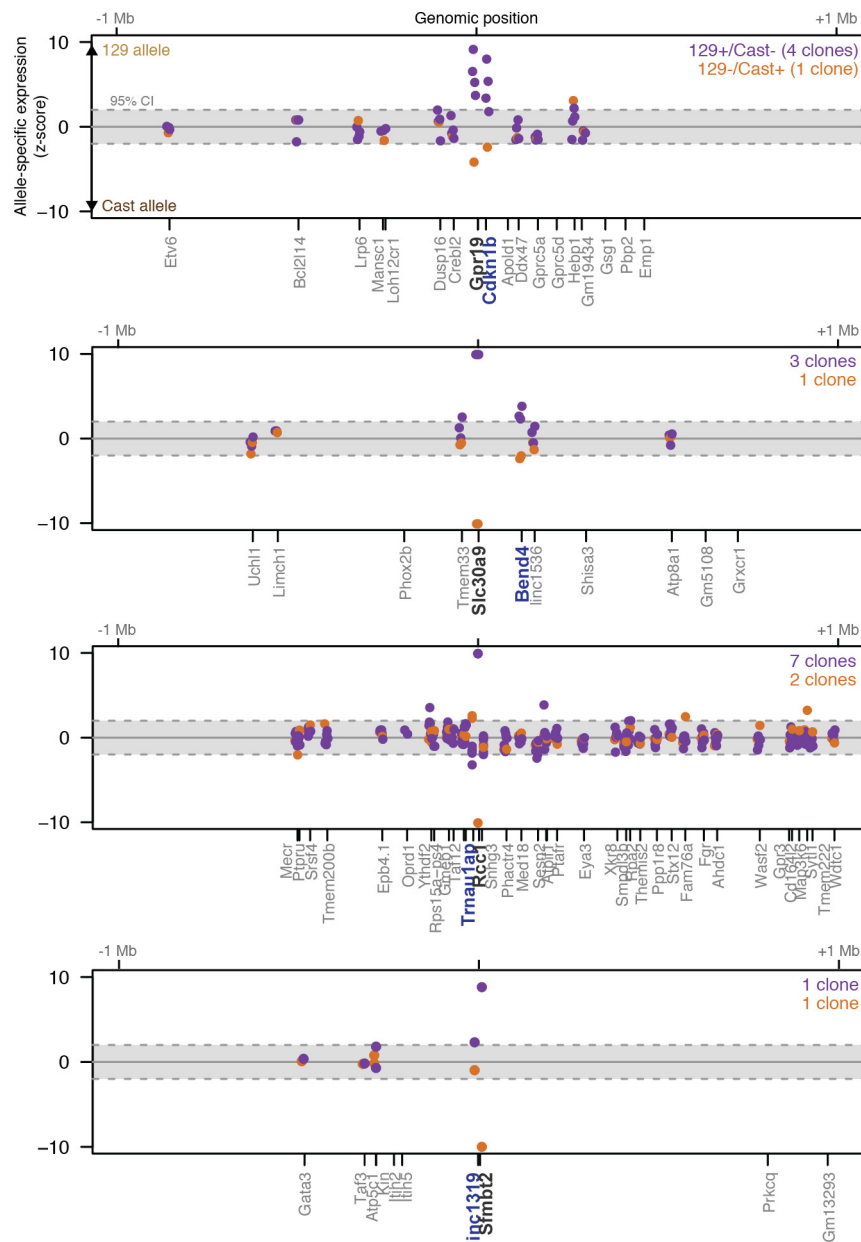
Extended Data Figure 3 | Read-through transcription at *Meg3* and *Snhg3* loci. **a**, *Snhg3* promoter knockout reduces the levels of *Rcc1* mRNA by 23%. However, sequencing of chromatin-associated RNA shows that transcription continues past the annotated 3' end of *Snhg3* into the downstream *Rcc1* gene (see Methods). This read-through transcription creates a fusion transcript containing exons of both *Snhg3* and *Rcc1*, as well as intergenic RNA. We note that this fusion transcript is also annotated in the syntenic human locus as an alternative isoform of *RCC1*. Bars, relative poly(A)⁺ RNA expression on modified versus unmodified

alleles. Error bars, 95% confidence interval for the mean ($n \geq 2$ alleles, see Supplementary Table 1). **b**, *Meg3* promoter knockout eliminates the expression not only of *Meg3* but also of two additional lncRNAs encoded downstream in a tandem orientation (*Rian* and *Mirg*). Although these three lncRNAs are annotated as separate genes, they appear to be derived from a single transcript driven by the *Meg3* promoter. This is consistent with the presence of continuous chromatin-associated RNA throughout the locus and a lack of CAGE reads at the 5' ends of *Rian* and *Mirg3*.



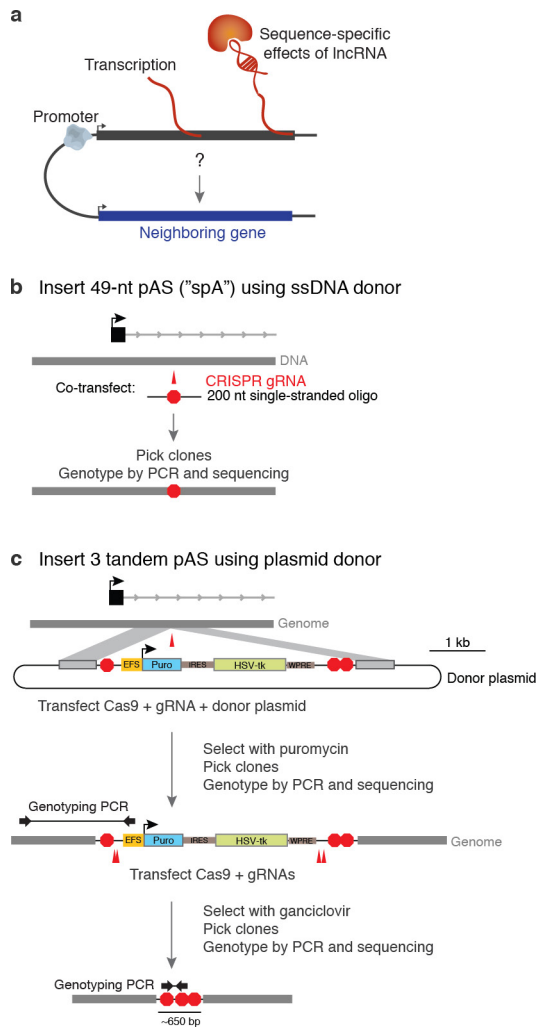
Extended Data Figure 4 | Promoter knockouts for five intergenic lncRNAs affect the expression of a neighbouring gene. Significance (z-score) of allele-specific expression ratios at all genes within 1 Mb of each of five lncRNA loci. Each dot represents a different heterozygous promoter knockout clone for a given gene. Dots are shown only for genes that are sufficiently highly expressed to assess allele-specific expression

(see Methods). The y axis is capped at -10 to +10 standard deviations from the mean. Black, knocked-out lncRNA; blue, gene with significant allele-specific change in gene expression (FDR < 10%). Independent clones are not expected to yield the same significance value (z-score), in part because read depth differs between samples.

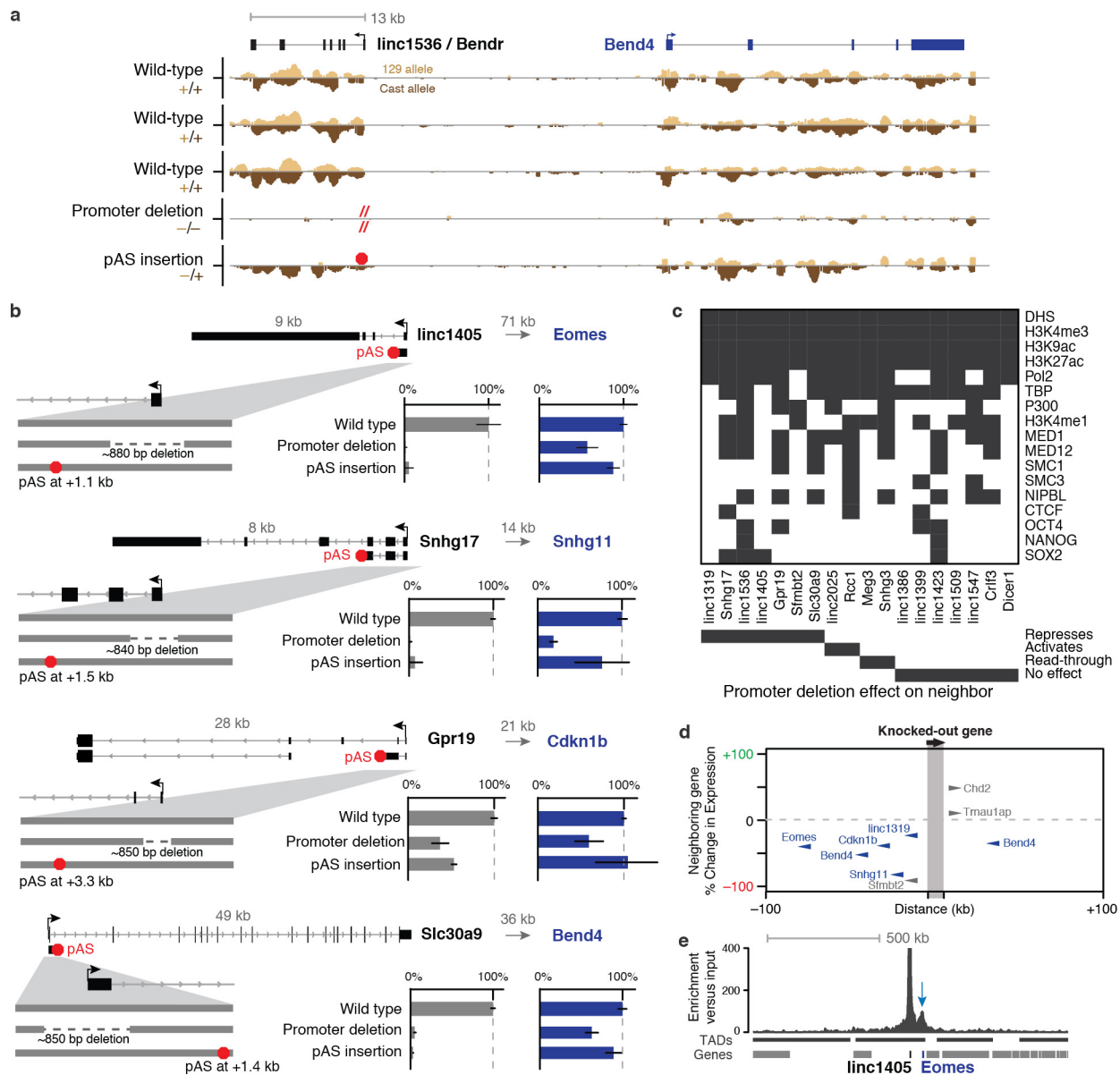


Extended Data Figure 5 | Promoter knockouts for four mRNAs affect the expression of a neighbouring gene. Significance (z-score) of allele-specific expression ratios at all genes within 1 Mb of each of four mRNA loci. Each dot represents a different heterozygous promoter knockout clone for a given gene. Dots are shown only for genes that are sufficiently highly expressed to assess allele-specific expression (see Methods).

The y axis is capped at -10 to +10 standard deviations from the mean. Black, knocked-out lncRNA; blue, gene with significant allele-specific change in gene expression (FDR < 10%). Independent clones are not expected to yield the same significance value (z-score), in part because read depth differs between samples.

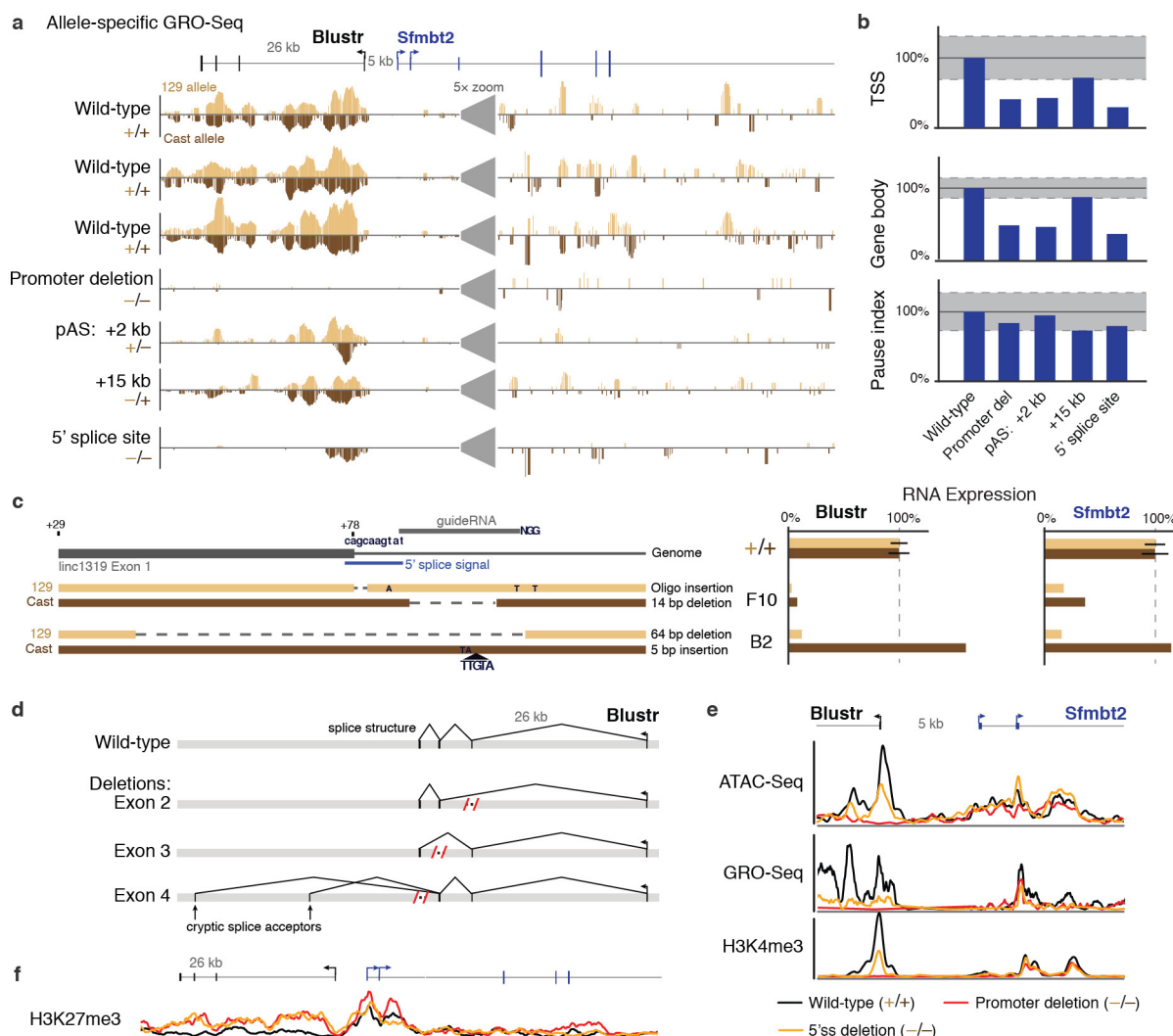


Extended Data Figure 6 | Dissecting mechanisms for how gene loci regulate a neighbour. **a**, Three categories of possible mechanisms by which a gene locus might regulate the expression of a neighbour. **b**, We used two strategies to insert pAS downstream of gene promoters. In the first strategy, we inserted a 49-bp synthetic pAS (spA) using a single-stranded DNA oligo with 75-bp homology arms (see Methods). **c**, In the second pAS insertion strategy, we cloned a donor plasmid containing a selection cassette and three different pAS sequences (see Methods). Homology arms of 300–800 bp were used to integrate the cassette. After isolating clones with successful insertions, we used a second round of transfections to remove the selection cassette, leaving behind three tandem pASs. EFS, elongation factor 1 promoter; Puro, puromycin resistance gene (*pac*); HSV-tk, herpes simplex virus thymidine kinase.



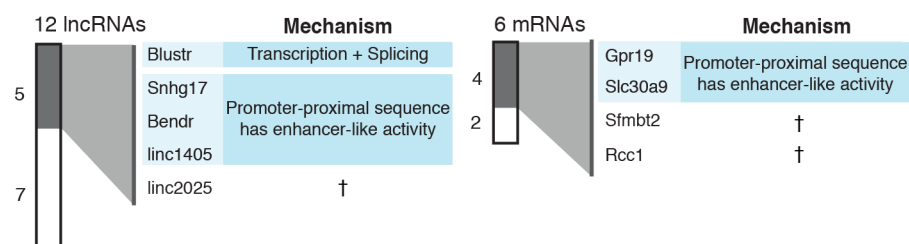
Extended Data Figure 7 | Promoters of lncRNAs and mRNAs have enhancer-like functions. **a**, Allele-specific GRO-seq signal for clones with the indicated modifications at the *Bendr* locus. Only reads specifically mapping to one of the two alleles are shown. The y axis scale represents normalized read count and is the same for all tracks. **b**, Allele-specific poly(A)⁺ RNA expression for genetic modifications at the *linc1405*, *Snhg17*, *Gpr19*, and *Slc30a9* loci. Bars, average RNA expression on modified compared to unmodified (wild-type) alleles. Error bars, 95% confidence intervals for the mean ($n \geq 2$ alleles, see Supplementary Table 1). Grey arrows indicate distance from the targeted locus promoter to the affected neighbouring gene. We note that, based on their location, the *Snhg17* and *Gpr19* pAS insertions probably allow more substantial splicing and transcription; for these loci, it is clear that the majority of the transcript is dispensable but it is possible that transcription close to

the promoter may be involved in the *cis*-regulatory function. **c**, Presence (grey) or absence (white) of various chromatin marks and transcription factors in mES cells in a 1.5-kb window centred on the TSS of each targeted gene. **d**, Distance from each knocked-out gene to its neighbouring target gene (x axis) versus the magnitude of the effect on the expression of the neighbouring gene (per cent compared to wild-type, y axis). Blue genes represent those discussed in main text; grey genes are discussed in Supplementary Note 5. **e**, Proximity-based contacts between the *linc1405* and *Eomes* loci. The y axis shows enrichment in a sequencing-based proximity assay in which we used antisense oligos to capture *linc1405* DNA and any interacting, cross-linked proximal DNA (see Methods). TAD annotations are derived from Hi-C experiments in mES cells (see Methods). Blue arrow, focal contact between the *linc1405* and *Eomes* loci.

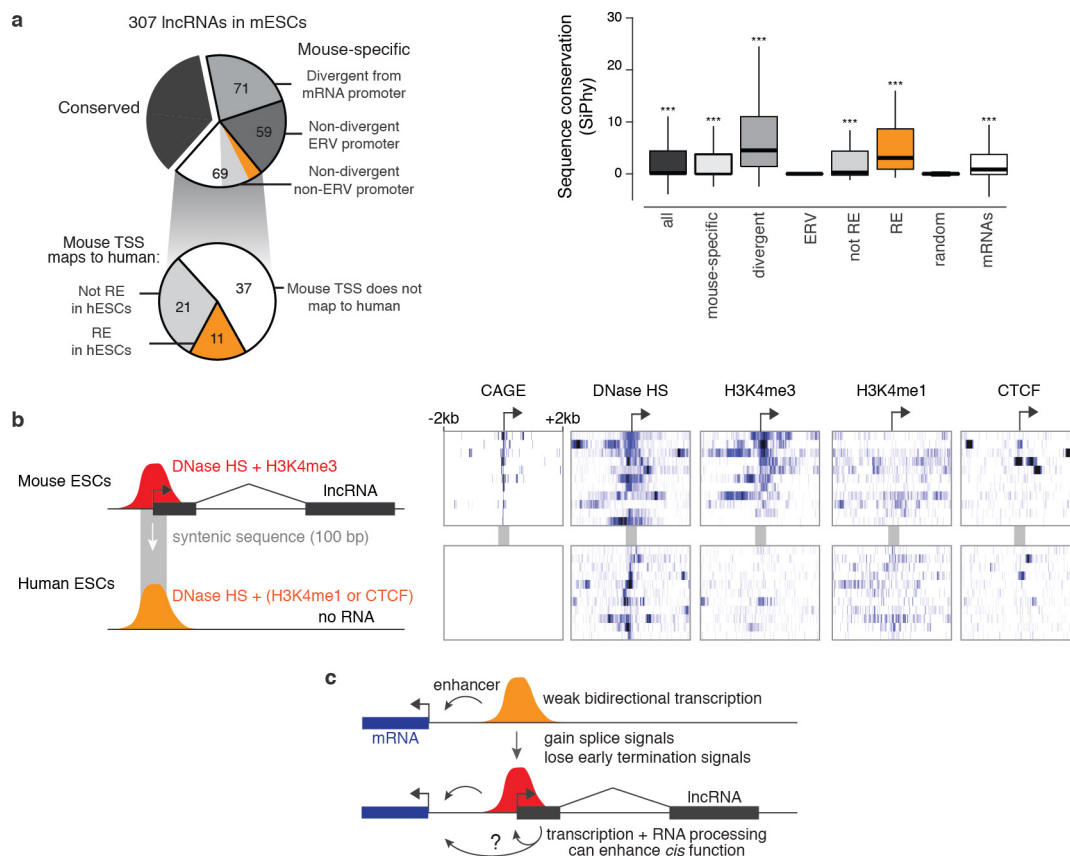


Extended Data Figure 8 | Characterization of genetic modifications in the *Blustr* locus. **a**, Allele-specific GRO-seq signal for clones with the indicated modifications at the *Blustr* locus. Only reads specifically mapping to one of the two alleles are shown. The y axis scale represents normalized read count and is the same for all tracks, and is magnified five times at the indicated location to better visualize the reads in the *Sfmbt2* locus. **b**, Quantification of allele-specific GRO-seq signal in the *Sfmbt2* locus on alleles modified as indicated. TSS, region including the two alternative TSSs of *Sfmbt2* and 2 kb downstream; gene body, region containing the remainder of the *Sfmbt2* gene locus; pause index, ratio of TSS to gene body. Dashed grey lines indicate the 95% confidence intervals for the mean of eight wild-type clones. Bars, $n = 8$ for wild-type and $n = 1$ for others. **c**, Schematic of the 5' end of the *Blustr* locus and genotypes of two knockout clones. The 5' splice site is located 78 bp downstream of the *Blustr* transcription start site (in this panel, *Blustr* is transcribed from left to right). One of the alleles from the two clones contains insertion of the oligo mediated by homologous recombination; the remaining three alleles contain insertions or deletions resulting from non-homologous end joining repair of sgRNA-mediated double-strand breaks, some of

which also disrupt the 5' splice site. Bar plots show allele-specific RNA expression for knockout clones and control clones ($n = 18$ for +/+, 1 for others). Error bars, 95% confidence interval for the mean. **d**, Schematic of the observed splice structures of *Blustr* RNA transcripts in poly(A)⁺ RNA sequencing of the exon deletion clones. Each deletion removes a region including ~50–200 bp on either side of the exon, thereby removing both the exon and its splice sites. The Exon 4 deletion removes the endogenous pAS, leading to new isoforms of the lncRNA transcript that splice into two cryptic splice acceptors downstream. **e**, GRO-seq, H3K4me3 ChIP-seq, and chromatin accessibility (ATAC-seq FPKM) at the *Blustr* and *Sfmbt2* promoters in cell lines with the indicated genotypes. Deletion of the first 5' splice site leads to a significant reduction in H3K4me3, RNA polymerase occupancy, and chromatin accessibility at the *Blustr* promoter, as well as H3K4me3 and RNA polymerase occupancy (but not accessibility) at the *Sfmbt2* promoter. **f**, H3K27me3 ChIP-seq at the *Blustr* and *Sfmbt2* loci in cell lines with the indicated genotypes. Deletion of the *Blustr* promoter or 5' splice site leads to spreading of the repression-associated H3K27me3 modification across a ~30 kb region.



Extended Data Figure 9 | Mechanisms for cross-talk between neighbouring lncRNAs and mRNAs. Proposed mechanisms based on pAS insertion experiments and other genetic manipulations (see text). For proposed mechanisms of lncRNAs marked with daggers see Supplementary Note 5.



Extended Data Figure 10 | Classification of lncRNAs based on conservation and promoter location. **a**, Classification of 307 lncRNAs expressed in mES cells. 'Conserved' transcripts are those that show significant evidence of cap analysis of gene expression (CAGE) data and/or poly(A)⁺ RNA in syntenic loci (see Methods). Divergent, initiating within 500 bp of an mRNA TSS, on the opposite strand; ERV, endogenous retroviral repetitive element (see Supplementary Note 9). Box plot shows sequence-level conservation of the promoters of subsets of lncRNAs expressed in mES cells. Random intergenic regions are matched to lncRNA promoters by GC content. Positive SiPhy score indicates evolutionary constraint on functional sequences. Orange category corresponds to mouse-specific lncRNAs that appear to have evolved from ancestral regulatory elements (REs) and correspond to sequences that show evidence for DNase I hypersensitivity in human embryonic stem cells. Significance is calculated compared to random intergenic regions using

a Mann-Whitney *U*-test. *** $P < 0.001$. Box represents first and third quartiles; centre line represents median; whiskers represent data within $1.5 \times$ the interquartile range. **b**, Chromatin and RNA data for 11 mouse-specific lncRNAs that appear to have evolved from ancestral regulatory elements. In mouse, these elements show evidence for CAGE, H3K4me3, and DNase I hypersensitivity, consistent with their roles as promoters. The syntenic sequences in human do not show evidence for CAGE but nonetheless are DNase I hypersensitive and are frequently marked by H3K4me1 and/or CTCF. **c**, Model for evolution of lncRNAs from pre-existing enhancers, which often initiate weak bidirectional transcription. Spliced transcripts may neutrally appear through the appearance of splice signals and loss of polyadenylation signals. In some cases, transcription, splicing, or other RNA processing mechanisms may feed back and contribute to the *cis*-regulatory function of the promoter, producing a lncRNA as a by-product.

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature19808

Corrigendum: An early geodynamo driven by exsolution of mantle components from Earth's core

James Badro, Julien Siebert & Francis Nimmo

Nature **536**, 326–328 (2016); doi:10.1038/nature18594

Some of the analyses (samples X1–2, X1–3, X1–4 and X2–4) reported in this Letter's Extended Data Table 1 and in the Supplementary Data spreadsheet were previously published in refs 28 and 30. The other analyses are new measurements, as are all measurements of aluminium concentration in the metal we studied. Owing to a typographical error, an erroneous value for the magnesium composition of sample X2–4 was reported in refs 28 and 30 (which has no impact on the conclusions of those papers because magnesium concentration was not used). Our recent re-evaluation of the raw microprobe data confirms that we have here reported the correct value (0.94 mol% in Extended Data Table 1, corresponding to 0.59 wt% in the Supplementary Data spreadsheet) for the magnesium composition of sample X2–4. (Therefore, no correction to this Letter is required.)

CAREERS

ON BALANCE Biotech post allows time for life **p.459**

BLOG Counsel and tales to help you reach your career goals blogs.nature.com/naturejobs

 **NATUREJOBS** For the latest career listings and advice www.naturejobs.com

ADAPTED FROM AKINDO/GETTY



EDUCATION

The gamble of a PhD hiatus

Pressure to barrel through a PhD can make taking a break seem daunting — but there are ways to limit career damage.

BY ERYN BROWN

Life can get in the way of science, forcing PhD students to take time out from the pursuit of knowledge and the lab.

It can be a tough call for students to make. Immersed in their work with no assurances of a great job, driven to scour the literature to stay current and primed to worry about competition and impressing their advisers, many PhD students think that academic success is everything. For them, nothing comes before their studies and research programme.

Breaks are risky — there is no way to ensure

a smooth return to studies, funding and the bench. University policies governing gap time vary across nations, regions and institutions, and maintaining funding and research continuity can pose hurdles. Attitudes towards time off also differ widely. Many faculty members and potential future employers look askance at a doctoral student's decision to step aside, even for a brief period.

Your capacity to put your PhD programme on hold will depend largely on your field, your institution and your advisers. In general, you can take a break when you need it, as long as you are prepared for the consequences

— particularly if you aim to pursue an academic career. The decision could affect your reputation, publishing record and ability to stay current with your research programme. But with careful planning, there are ways to soften the blow (see ‘How to take a successful break’).

Few statistics exist on how often, for how long or why PhD students take time off from their studies. In the United States, neither the National Science Foundation nor the Council of Graduate Schools tracks leaves of absence or can point to a central source for such data.

Some individual institutions provide estimates of how many PhD students have ►

► taken breaks each year. Heather Amos, a spokesperson for the University of British Columbia in Vancouver, Canada, says that about 50 PhD students out of nearly 4,000 across all disciplines, or about 1.25%, took a leave of absence in 2015.

Martin Grund, spokesperson for the Max Planck Institutes' graduate-student organization, PhDnet, says that his group doesn't track leaves of absence.

But, he says, internal surveys show that 7% of doctoral students at the institutes in 2012 were parents, and so had probably taken parental leave at some point.

"The way technology is changing, taking a break can become difficult. How do you stay cutting edge?"

Some funding agencies allow for certain interruptions of study, including care for children and elderly people, professional development and other life needs. Some universities permit students to retain access to campus services while on leave for a variety of reasons; others have no defined policy.

Anecdotally, it seems that few PhD students so much as think about a pause in their programme. "I think most don't even consider it," says Heather Buschman, who earned a PhD in molecular pathology from the University of California, San Diego, School of Medicine after taking six months off for a US National Cancer Institute communications internship in 2006. "They think, 'I could never do that. People are on such a focused trajectory and see any wavering as a negative.'"

There is a great deal of external and internal pressure to race to the finish, agrees Gareth O'Neill, a PhD candidate in linguistics at Leiden University in the Netherlands. "Doing a PhD is a relatively focused and driven occupation — once started, you just want to finish it," he says.

As a board member of the PhD Candidates' Network of the Netherlands, O'Neill is involved with an initiative called the Professional PhD Program, which helps to place PhD students who seek work experience outside academia. O'Neill says that the programme rarely receives applications from students who feel they need to stay at the bench throughout their doctoral studies, but that those who do apply sometimes experience pressure from supervisors to finish their PhD sooner. That pressure, he adds, is misguided or inapplicable — particularly from mid- or late-career scholars, who don't know or who don't want to admit how hard it is for new PhD students to remain in academia now. "We hope to bring about a shift of mindset," he says.

HIGHTAIL IT OUT

Still, when the need for a hiatus arises, some don't hesitate to take it — and then sail through their leave and back. Earlier this

year, Anna Miller earned a PhD in parks, recreation and tourism management from North Carolina State University (NCSU) in Raleigh. She says that she never questioned her decision to leave her research behind for half a year, when her Brazilian fiancé was offered a postdoctoral appointment in Portugal. The travails of a long-distance relationship had become burdensome, and she wanted to join him abroad. "Academically, I was getting a bit burnt out, but it was really the strain on my personal life that was the problem," she says. "If I was going to stay in the programme, I needed to deal with the personal part of my life."

Her adviser was concerned that she might not come back, and Miller herself says that she left for Portugal knowing that might be true. But in Lisbon, she found herself drawn to nearby parks, and started studying how they were managed, just for fun.

"It was a refreshing way to look at the same questions from a different perspective and reaffirm my desire to study this subject," she says. "I came back with new energy for being a full-time student."

Re-entry turned out to be easy. To get approval for a leave of absence, Miller and her advisers had already agreed on a formal plan for her return, charting out how she would later complete course work, research and exams.

They had also predetermined how Miller's funding, which was suspended while she was away, would be reinstated. Everything unfolded as planned — and Miller became treasurer and then co-president of her department's graduate-student association. She also began to mentor other students and to organize career panels and other programmes.

Three years on, Miller and her fiancé have since married, and she is now a resident lecturer at the School for Field Studies Center for Marine Resource Studies in the Turks and Caicos Islands. She teaches undergraduates

who are studying abroad. "I can't think of any negatives of taking the time off," she says.

Others also report a positive experience. "Ideally, I'd say don't take time off, but if you do, don't judge yourself harshly," says Jen O'Keefe, a geologist and science-education researcher at Morehead State University in Kentucky.

She took a pause from her PhD studies in 2002, after a working relationship with an adviser fell apart. Ultimately, she devised a new plan that combined part-time work on her doctorate with a full-time teaching schedule.

Looking back, she thinks that her research career benefited from the five-month break, which enabled her to refocus her work towards palaeoecology and curriculum and instruction, as well as a variety of other pursuits that she loves. "Everything from fly-ash geochemistry to honey studies to sinkholes," she says. "No two PhD situations are the same. You have to do what's right for you."

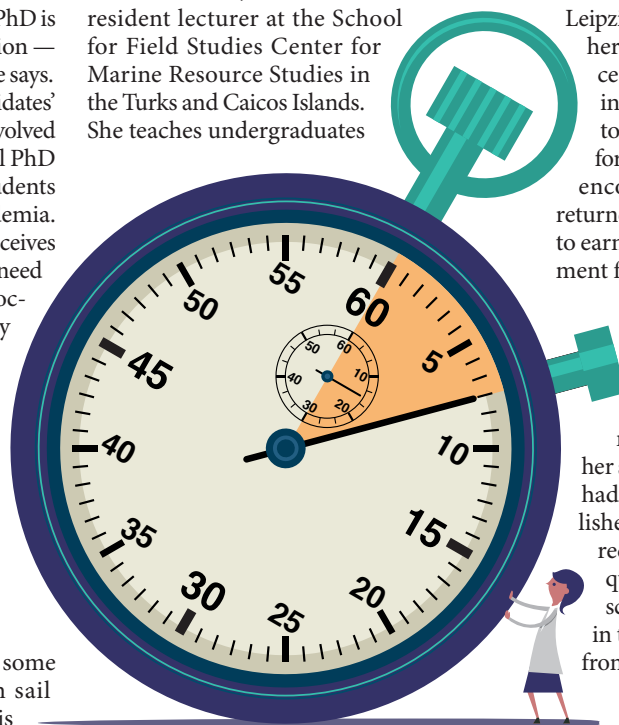
Some think that their field of study smoothed the way. Benedikt Herwerth, who studies theoretical quantum physics at the Max Planck Institute of Quantum Optics in Garching, Germany, says that he had little trouble setting up two stints of paternity leave, for a total of seven months, after the birth of his daughter in February. He says that Germany's generous approach towards parental leave helped, but that his field of study might also have facilitated the interruption. "I'm not doing experiments," he says. "It might be an advantage."

PERSONAL BARRIERS

Even when there are no obstacles to taking time off, trouble might arise that complicates a student's return. Eleanor Harding, a doctoral candidate at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany, has taken two breaks from her PhD research on how the brain processes music and language. The first was in 2011, when her mother died. Harding took several months off. "I lost my edge for quite a while," she says. "But my adviser encouraged me to keep going." Harding returned to work later that year and expected to earn her degree in 2013 — until an experiment fell through, which caused delays, and pregnancy complications rendered her unable to work.

In May 2014, after her daughter was born, Harding returned to her research, but she found that she could not afford enough childcare to resume her studies full-time. In addition, while she had been out, other researchers had published work in her area, so she had to redirect her research to examine a narrower question that would respond to the other scientists' work. "Now, unfortunately, I'm in the middle of the pack instead of at the front," she says. "I can't say I was the first."

A gap of just one year can put a PhD candidate behind when it comes to



ADAPTED FROM SORBETTO/GETTY

DAMAGE CONTROL

How to take a successful break

Even when you know that stepping away from your PhD studies makes sense, it can create problems, from disruptions in your research to unmerited damage to your reputation. To finish your programme and launch your career, try to take these steps.

● **Plan, plan, plan** PhD candidates who successfully take leave say that charting out their eventual return was as important as departing in the first place. Make sure that you understand how your time away will affect your funding, teaching responsibilities or other roles in your department. Plan how you'll continue your research once you return, even if you think its specific direction could change.

● **Find your champion** Make sure you have a person in your corner — ideally, your adviser — to stick up for you when you encounter resistance. “You have to have someone who would say, ‘This person is worth carrying for six months, or a year, because they bring great value to my lab and to the institution,’” says Nathan

Vanderford, who researches training issues in academia at the University of Kentucky in Lexington. “You have to have an advocate.” Others, including savvy administrative staff, can prove key in the nuts and bolts of making arrangements.

● **Open your mind** One reason that students decide not to take a leave of absence is because of fear that it will damage their chances of getting a conventional academic job. But a little experience outside the ivory tower might teach you that you would actually be happier applying your degree elsewhere — in industry, for instance. “All the things I really liked the most were the things most scientists hated,” says Heather Buschman, who earned her PhD from the University of California, San Diego. In Buschman's case, having an adviser who had worked as a clinician and encouraged her to pursue a different direction — communications — was crucial, she says. **E.B.**

mastery of important technological advances, warns Kim First, president and chief executive of the recruiting firm Agency Worldwide in Encino, California. As a headhunter who searches for PhD graduates for jobs in biotechnology and pharmaceutical companies, she says that she encounters few candidates who have interrupted their doctoral programme. “The way technology is changing, taking a break can become difficult,” she says. “How do you stay cutting edge?”

Other recruiters say that taking time away to have children or for other life events can hurt a researcher's scientific reputation, and that students should find ways to incorporate those obligations into their PhD programme without putting their research on pause.

Some think that the stigma might be worse for women. Justin Schwartz, head of the materials science and engineering department at NCSU, has helped students to organize leaves of absence. When it comes to parental leave, he says, women are more likely than men to take the time off — but those who do are often terrified (sadly, with some reason, he notes) that faculty members will think that they lack the drive to be the best and will extrapolate that women aren't suited to doing science.

But whether female or male, most students experience one clear consequence after taking the break: they lose momentum. Harding says that although there was a benefit to delaying her dissertation — a competing paper helped her to solve a problem in her data — she now has few job leads near

her husband's medical residency in the Netherlands, and attributes that to having lost potential publications and chances to attend more conferences.

“Your worth is based on quantitative measures like an impact factor,” she says. “They want people with publications. Life doesn't always cooperate.” Harding is now networking locally — getting involved, for instance, with a organization in the region that funds research into Parkinson's disease.

O'Keefe wishes that the harsh judgement weren't there, but says that it seems specific to academia. “People feel badly and a lot of scientists out there judge them harshly,” she says. “There's a lot of, ‘If you had to take time off, you're not really good enough to finish.’” She says that many early-career scientists she knows who interrupted their PhD programmes eschewed academic research in the end, and instead, accepted positions in industry or teaching. Now in her early 40s and a mother, she says that she wouldn't have done anything differently, and looks forward to expanding her research.

“I was on the fast track and I was moving too fast,” she says. “A lot of good comes from taking a break and reassessing your priorities. A year off is sometimes the best thing you can do. The big message is, it's OK and you're not alone and you can go on to be what you want to be.” ■

Eryn Brown is a freelance writer in Los Angeles, California.

TRADE TALK
Industrial shift

As a staff scientist at Thermo Fisher Scientific in San Diego, California, Shikha Mishra investigates how to deliver DNA and other materials into cells and organisms. She explains how

and why she left academia for a job in industrial research.

Describe your workday yesterday.

I had run an experiment the night before and had a bunch of data. I analysed them for about an hour and talked with my manager about what I'd seen and tried. Then we formulated a plan and I went back into the lab and set up another experiment. That was pretty standard — very similar to what goes on in academia.

How did you learn about jobs outside academia?

My father was an industrial physicist, and I knew that he was happy. But my biggest difficulty was that my friends and I didn't understand where we could fit within the biotech industry. I didn't know the different sectors: manufacturing and quality control, marketing, product management and development. The way I learned was buying many, many cups of coffee for working people who agreed to meet with me. I'd ask: “What do you do? How did you find this opportunity?”

When did you decide on industry?

Four years into my postdoc at Harvard Medical School in Boston, Massachusetts, I was happy and my projects were going well — but the commitment time-wise was extraordinary. It was 7 days a week, and there were times when it was 14 or 16 hours of work a day. I was coming to a transition because I was about to get married. I thought: “Life has got to fit into my career plans.”

What do you wish you had done differently in your job search?

I wasn't thinking past the grant deadline or the next publication. I should have had a LinkedIn profile before starting my job search, because I had to track down old friends — contacts whom I may not talk to every day, but who'd be happy to help me. ■

INTERVIEW BY MONYA BAKER

This interview has been edited for length and clarity. For more, see go.nature.com/2fwrzyg

BOX 27

A question of definition.

BY KEVIN LAUDERDALE

“Welcome to Galactic Confederation processing. Please state your name.”

The creature behind the counter — they used actual counters, even this deep in the heart of the Galaxy — looked like a giant chartreuse bear with black fins running down her spine. Marta assumed the bear was a she. The translator certainly produced a female voice. The bear wasn't wearing a name tag or anything.

“Ambassador Marta Rilla-Chen,” Marta replied clearly and directly. “Of the Earth Diplomatic Corps.” She was there to present her credentials to the Confederation. But the ceremony wasn't being held in a throne room or a hall bedecked with ceremonial splendour. She was in an office of some sort, whose multiple counters and flickering lighting reminded her of the fluorescents in the DMV back home. Only the view of the swirling Galactic core out of the windows and the presence of dozens of different types of aliens around her broke the illusion. Still, she couldn't help but wonder if some of those other aliens in the office were there to pay parking tickets.

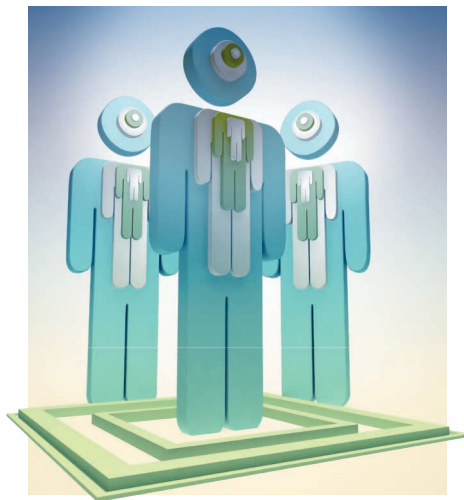
“Form, please,” growled the bear in a bored monotone, and Marta handed over the flexiplast pages. The bear ran her eyes down the form. She stamped an item here and circled an item there with a laser pen. She stopped and looked up from the form. With pronounced disapproval she said: “You didn't fill in Box 27.”

Marta had been expecting that. “Right,” she said. “I wasn't sure what to put down. What does ‘Society Code’ mean?”

The bear sighed. “Every member species has a specific number and name.” Marta found the droning quality of her rough voice actually quite charming. The bear had given this speech, what, a thousand times before? That's how many member worlds were in the Confederation. “We'll assign your catalogue number, but you need to provide the name you will all be known by.”

Marta felt confused. Her face must have shown it.

The bear said: “I can see we'll have to go through the whole *zarfasnop*. Let's try to make this quick. Says here your planet's name is Earth. Does that mean anything in particular?”



“Ummm, land, soil, dirt.”

“Oh yeah. I've heard about your planet. Earth is the place with all the water, but you named it Dirt. Makes a lot of sense. How about if you call yourselves The Crazies. That's short at least. Raljevanites call themselves We Who Created The Ice-Based Food That Is Sweet But Causes Your Head To Hurt. That's their planet's claim to fame and they're sticking with it for all it's worth. If you were from Melton II, you would be part of The Survivors. That's nice. Also short. Anyway, The Crazies, then?” Her pen hovered over the form.

“Wait!” cried Marta. *The Crazies!* She could just hear her boss exploding over that. Forget her boss — the 8 billion people of Earth would not be pleased that the rest of the Galaxy would know them as The Crazies. “Let's try something else. Ummm, we also call ourselves Terrans.”

“What does that mean?”

“People who live on Terra ... Oh, but that just means land.”

“Which is dirt. The Dirties?”

Marta shook her head. “Oh! Of course! We call ourselves People.” Why hadn't she thought of that in the first place? That was better. That was respectable.

The bear snorted. “Everyone calls themselves People. My people call themselves People. That guy over there from Aquatox IV wearing the glass bowl that keeps his head immersed in water, calls his fellow creatures People. You'll have to try harder.”

Marta thought. “All right. We are ... *Homo sapiens*. The Wise Man. And Woman. Technically we're *Homo sapiens sapiens*. So how about The Really Wise People?”

“Didn't your planet almost blow itself up with nuclear weapons —”

“But we didn't.”

“— and then almost destroy itself through overheating?”

“But we didn't.” Still, she could see the bear's point. Humans had eventually mastered atomic power constructively, restored the planet's ecology, and made it to the stars. Now they were even joining a confederation of fellow space-farers. We had come far, but perhaps it was stretching it a bit to wish to be known as Really Wise.

“Besides,” said the bear. “The Really Wise People is already taken.”

“By whom?” Marta was indignant. Who could be so egotistical as to actually use that?

The bear gave a wry, growling chuckle. “My People.”

Marta shook her long black hair and composed herself. “All right ... oh! The Humans!”

“What does that mean?”

Marta sighed. “People.”

“Your species has a real talent for circular thought. You want to reconsider The Crazies?”

“No.” Marta closed her eyes and concentrated hard. Humans were indeed crazy, dirty, noisy, lovable, brilliant, joyous ... Marta wasn't getting anywhere. This was frustrating. She looked at the bear. “What's your name?” She couldn't keep on thinking of this Really Wise Person as a bear.

“Why do you want to know?”

“It's what humans do. We connect.”

“I am called Weylo.”

“Hi, Weylo.” Marta smiled. “Truth is, we humans don't really know who we are. We're just beginning to figure that out. That's partly why we're out here. Each of us is a human, but we're really at our best when we're connected. It's the things we do together that makes us Humanity.”

“Humanity?”

“Yeah, Humanity.” That was it. It meant not just all of us, but all of us connected.

“At least it's short.” Was that a smile on Weylo's muzzle? She gave the form one final, resounding stamp. “Welcome to the Galactic Confederation, Marta Rilla-Chen ... and Humanity.” ■

Kevin Lauderdale's fiction has appeared in several of Pocket Books' Star Trek anthologies as well as numerous cross-genre collections. Visit him at KevinLauderdale.Livejournal.com.

ILLUSTRATION BY JACEY

➔ NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm